

180  
#27  
183  
12-1-95



**IN THE UNITED STATES PATENT AND TRADEMARK OFFICE**  
**BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES**

In re Application of:  
Adams *et al.*

Serial No. 08/113,561

**Filed: August 25, 1993**

**For: METHODS AND COMPOSITIONS  
FOR THE PRODUCTION OF  
STABLY TRANSFORMED,  
FERTILE MONOCOT PLANTS  
AND CELLS THEREOF**

**Group Art Unit: 1804**

**Examiner: G. Benzion**

**Atty. Dkt.: DEKM:055/PAR**

**CERTIFICATE OF MAILING**  
**37 C.F.R. 1.8**

I hereby certify that this correspondence is being deposited with the U.S. Postal Service as First Class Mail in an envelope addressed to: Assistant Commissioner for Patents, Washington, D.C. 20231, on the date below:

October 18, 1995

Date David L. Parker

## BRIEF ON APPEAL

RECEIVED

NOV 21 1995

Assistant Commissioner for Patents  
Washington, D.C. 20231

GROUP 1800

**Sir:**

This Brief is filed in response to the final Office Action mailed January 31, 1995, regarding the above-captioned application. This brief is due July 31, 1995, by virtue of the Notice of Appeal filed on May 31, 1995.

7-2-1 10/25 081235-1

**I. STATUS OF THE CLAIMS**

Claims 1-67 were filed with the application, and various claims have been withdrawn, and various claims amended, during the course of prosecution. Claims 2-4, 47 and 50-58 and 60-67 are currently pending and are the subject of the present appeal. A copy of these pending claims is attached as Exhibit A.

**II. STATUS OF THE AMENDMENTS**

An amendment after final was sought in an amendment filed on June 12, 1995, and these amendments were introduced per an Advisory Action dated July 27, 1995.

A further amendment is sought concurrently with the filing of the present Brief. It is believed that these amendments should be acceptable to the Examiner as only claims and species have been removed, and no new issues presented.

**III. REAL PARTIES IN INTEREST**

The real party in interest in the present appeal is DeKalb Genetics Corporation, assignee of the captioned application.

**IV. RELATED APPEALS AND INTERFERENCES**

An appeal is currently pending in a related case, USSN 07/565,844.

A notice of appeal has been filed in USSN 08/112,245.

## V. SUMMARY OF THE INVENTION

The invention of the claims pending in this Appeal relates to fertile, transgenic maize (corn) plants having a selected, specified "transgene" introduced into their genetic makeup. There is no single claim pending that is generic to all of the pending claims, and the claims are instead directed to a large number of what are submitted to be patentably distinct species. In particular, the independent claims on appeal are directed to fertile, transgenic corn plants having one or more of the following genes inserted into their genome:

### Claim 47 -- Selectable or Screenable Marker Genes

- an aequorin gene
- a gene encoding a cell wall protein
- an HPRG gene

### Claim 60 -- Negatively Selectable Marker Genes

- a cytosine deaminase gene
- a T-DNA gene 2
- an antisense *bar* gene
- an antisense *nptII* gene

### Claim 61 -- Inducible or Tissue Specific Promoter or Enhancers

- an  $\alpha$ -tubulin promoter
- an *ocs* promoter
- an ABA-inducible promoter
- a turgor-inducible promoter

### Claim 62 -- Herbicide Resistance Genes

- a *bxn* gene

### Claim 63 -- Insect Resistance Genes

- an oryzacystatin gene

a wheat or barley amylase inhibitor gene  
a lipoxygenase gene  
an ecdysteroid UDP-glucosyl transferase gene  
a DIMBOA synthetic gene of the *bx* locus

Claim 64 -- Disease Resistance Genes

a pathogenesis related (PR) protein gene

Claim 65 -- Stress Resistance Genes

a glycerol-3-phosphate acetyltransferase gene  
a superoxide dismutase gene  
a glutathione reductase gene

Claim 66 -- Drought Resistance Genes

a mannitol-1-phosphate dehydrogenase gene  
a trehalose-6-phosphate synthase gene  
a myoinositol 0-methyltransferase gene  
a Late Embryogenic Protein (LEA) gene

Claim 67 -- Grain Composition Genes

an acetyl-CoA carboxylase gene  
an ACP-acyltransferase gene  
a b-ketoacyl-ACP synthase gene  
an acyl carrier protein gene  
a fatty acid desaturase gene  
a fatty acid epoxidase gene  
a fatty acid hydratase gene  
a fatty acid dehydratase gene  
a sense or antisense phytoene synthase gene  
a sense or antisense phytoene desaturase gene  
a sense or antisense lycopene synthase gene  
a phytase gene  
an ADP-glucose pyrophosphorylase gene  
a starch synthase gene  
a starch branching enzyme gene  
a sucrose synthase gene



The remaining claims specify cells, seed and progeny, particular preferred promoters or other control elements, or breakout individual species.

It is Appellants' position that the more than 45 different species of corn plants represented by the above independent claims on appeal are separate and patentably distinct.

## **VI. ISSUES ON APPEAL**

It is believed that the following three rejections are the only issues that remain for appeal:

- (1) The rejection of all the claims on the basis of provisional obviousness-type double patenting over copending USSN 07/508,045, filed April 11, 1990 (the '045 specification, attached as Exhibit B);
- (2) The rejection of all of the claims as obvious over the '045 application; and
- (3) The rejection of all of the claims as obvious over the Goldman et al. patent (hereinafter "Goldman"), US 5,187,073 (Exhibit C) alone or in combination with what the Examiner states are Appellants' acknowledged state of the art (see Office Action of 6/28/94, page 7).

## **VII. GROUPING OF THE CLAIMS**

The claims will stand or fall separately. It is Appellants position that a fertile, transgenic maize plant bearing one of the recited genetic elements is patentably distinct from a second transgenic maize plant bearing another of the recited genetic elements. The subject matter of each claim must be judged in accord with the applicable standards of obviousness, and the claims are patentably distinct absent proof of obviousness. Even then, the claimed corn plants with

distinct genes would be considered obvious only where secondary indicia of non-obviousness are absent.

### **VIII. ARGUMENT**

#### **A. PROVISIONAL REJECTION OF CLAIMS ON THE BASIS OF OBVIOUSNESS-TYPE DOUBLE PATENTING OVER THE CLAIMS OF THE '045 APPLICATION**

##### ***SUMMARY OF THE REJECTION***

The Action first rejects all of the claims on the basis of obviousness-type double patenting over claims 28-68 of the '045 application (see Exhibit B). The Examiner takes the position that the present claims are directed to genes and genetic elements previously known in the art, and that the introduction of these genes into maize (corn) is obvious in light of the claims of the '045 application.

##### ***SUMMARY OF APPELLANTS' ARGUMENT***

In response, it is respectfully submitted that the subject matter of the claims pending in the present appeal is distinct from the subject matter of the '045 application. The claims on appeal are directed to transgenic maize bearing particular genes that are novel and nonobvious in the context of fertile, transgenic corn plants: There is submitted to be no disclosures in the '045 specification that teaches, suggests or in any way motivates one of skill in the art to produce the particular transgenic corn plants of the pending claims, and no evidence in the prior art that the specified genetic elements would function in corn to achieve a useful result.

The Action fails to base its finding upon a teaching, suggestion or motivation to employ each of the claimed genes -- a prerequisite to a finding of obviousness-type double patenting. *Carmen Industries, Inc. v. Wahl*, 220 U.S.P.Q. 481, 487 (Fed. Cir. 1983); *Mirafi Inc. v. Murphy*, 14 U.S.P.Q. 1337, 1347 (N.C. 1989). To maintain the rejection it must be shown that the subject matter of the later claims are obvious over the claims of the earlier application. It is submitted that no such demonstration has been made here, and thus no *prima facie* rejection has been made.

***THE EXAMINER HAS NOT MADE OUT A PRIMA FACIE CASE OF  
OBVIOUSNESS-TYPE DOUBLE PATENTING***

A rejection on the basis of obviousness-type double patenting is treated in much the same way as obviousness rejections under 35 U.S.C. § 103. In order to make out a *prima facie* rejection, an examiner must demonstrate that subject claims of the later patent are obvious in light of the claims of the earlier patent. *In re Stanley*, 102 U.S.P.Q. 234 (CCPA 1954). Moreover, where the earlier patent discloses a genus and certain species, and the subject claims are directed to species not specifically disclosed in the earlier patent, the claims of the subject patent are patentably distinct absent evidence that the specific later species are obvious. *In re Sarett*, 140 U.S.P.Q. 474 (CCPA 1964).

**1. An Earlier Generic Patent Does Not Render *prima facie* Obvious Claims to a Later Species Under Double Patenting**

The proposition that an earlier generic patent and a later species/improvement patent does not raise obviousness-type double patenting concerns has been more recently reconfirmed by the Federal Circuit in the case of *In re Kaplan*, 229 U.S.P.Q. 678 (Fed. Cir. 1986), the court citing to E. Stringham's article *Double Patenting*:

One of the simplest, clearest, soundest and most essential principles of patent law, is that later invention may be validly patented, altho [sic] dominated by an earlier patent, whether to the same or to a different inventor.

229 U.S.P.Q. at 682.

Thus, the fact that the '045 claims may dominate the subject claims is totally irrelevant to whether the present claims are properly subject to an obviousness-type double patenting rejection. The fact remains: the examiner must make out a *prima facie* case that the present claims are obvious in light of the claims of the earlier application. It is submitted that the Examiner has failed to make out a proper *prima facie* case of obviousness.

The present claims are said to be obvious over those of the earlier-filed application (or claims) based on the fact that the earlier application was directed generically to "fertile, transgenic corn" transformed with any selected gene and the present application is directed to fertile, transgenic corn comprising one of 50 or so selected specific genes not mentioned in the '045 case. While it is true that the genes were known to exist *per se*, Appellants are unaware that these genes have been successfully placed into corn and there is no art relied upon by the Examiner to teach or suggest fertile, transgenic corn genetically engineered to express the specified genes.

**2. A Prior Art Genus Does Not Render *prima facie* Obvious a Later Species**

The examiner is relying merely upon a prior genus to obviate presently claimed species. It is submitted that obviousness of the present claims in light of the prior genus is governed by the Federal Circuit's holdings in *In re Jones*, 21 U.S.P.Q.2d 1941 (Fed. Cir. 1992) and *In re Baird*, 29 U.S.P.Q.2d 1550 (Fed. Cir. 1994). These cases held that a prior art genus does not render *prima facie* obvious a later species within that genus.

In *Jones*, the PTO Solicitor argued the prevailing PTO view that a prior art genus rendered *prima facie* obvious a later claim to a species within that genus. It was at that time routine practice in the PTO examining corps to base *prima facie* obviousness rejections on the existence of a prior art genus, thus shifting the burden to applicants to come forward with evidence of non-obviousness. The claim in *Jones* was directed to a new salt of the previously known herbicide "dicamba." The Solicitor argued that since "dicamba" *per se* was known in the art, and that the particular salt cation had been previously known, that the dicamba salt formed with the known cation was *prima facie* obvious. The court disagreed, and held that no *prima facie* rejection had been made, relying on the large number of species embraced by the prior art genus and the absence of any specific motivation to use the particular salt cation in the context of dicamba. 21 U.S.P.Q.2d at 1943.

The *Jones* court specifically rejected the Solicitor's argument that one of skill in the art would have been motivated to use the particular salt cation with dicamba since structurally similar salts had previously been prepared and were in the prior art. In rejecting the Solicitor's argument, the court observed that:

Before the PTO may combine the disclosures of two or more prior art references in order to establish *prima facie* obviousness, there must be some suggestion for doing so, found either in the references themselves or in the knowledge generally available to one of ordinary skill . . .

Conspicuously missing from this record is any *evidence*, other than the PTO's speculation (if it be called evidence) that one of ordinary skill in the herbicidal art would be motivated to make the modifications in the prior art salt necessary to arrive at the [claimed] salt. (underlining added)

21 U.S.P.Q. at 1943-44.

The Federal Circuit has recently reaffirmed the *Jones* holding in the case of *In re Baird*, 29 U.S.P.Q.2d 1550 (Fed. Cir. 1994). The *Baird* court followed *Jones*, stating that "[a] disclosure of millions of compounds does not render obvious a claim to three compounds". 29 U.S.P.Q.2d at 1552. The court again considered it particularly relevant that the prior art genus was quite broad, as in the present case, and focused on the fact that there was no specific motivation in the evidence or record for preparing the particular species of the subject claims.<sup>1</sup>

These facts of *Jones* are consistent with those of the present case. The prior '045 claims are generic claims (and some distinct species claims) directed fertile, transgenic plants. In contrast, the present claims are directed to specific corn plants bearing specific distinct introduced genes. Just as in *Jones*, it is agreed that the genes one would use to prepare the presently claimed transgenic corn species were known. However, also just as in *Jones*, the Examiner here has presented no evidence to demonstrate that one of skill in the art would be

---

<sup>1</sup> It is of interest to note that in March 1994 following the decision in *Baird*, Commissioner of Patents Lehman distributed a directive to examiners, instructing them to disregard the *Baird* decision in making *prima facie* rejections of species embraced by prior generic disclosures. See 1161 OG 314 (4/14/94). However, in an April 17, 1995 notice to examiners, the Commissioner rescinded the earlier notice, agreeing that *Baird* was good law. See 50 PTCJ 3 (May 4, 1995).

motivated to introduce the particular genes set forth in the claims into corn, and no evidence that one of skill would have predicted, when the present application was filed, that such genes would function appropriately in the resultant engineered corn.

***THE EXAMINER HAS NOT MET THE STANDARD OF JONES AND BAIRD***

The Examiner's only statement on the record regarding this obviousness-type double patenting rejection is set forth in the Office Actions of 6/28/94 and 4/17/95. In the 6/28/94 Action where the rejection was first entered, the Examiner stated merely that:

Although the conflicting claims are not identical, they are not patentably distinct from each other because each application is drawn to fertile transgenic maize in which the genome is augmented by the addition of DNA not normally found in maize or if found in maize is inserted, modified, altered or otherwise manipulated to the extent that it afford/s a change that is detectable over maize not transformed by phenotypic or genotypic change. Each application differs by the specific recitation of the DNA of interest which is deemed to be experimenter choice.

It is submitted that such a statement, with nothing more, fails to raise a *prima facie* case of obviousness-type double patenting under the standards set forth in *Sarett, Jones and Baird*. The mere recitation of a prior art genus does not meet these standards.

The Final Action of 4/17/95 observes that certain genetic starting materials were known in the art at the time the present invention was made. The Action argues that if the prior art genetic elements function in a predictable manner when placed into corn, then a *prima facie* case of obviousness has been shown. The Action has confused the applicable standards: The issue is not whether the prior art genetic elements actually function in corn, the issue is whether the prior art provides a proper motivation and evidences in advance that the invention would be successful.

Appellants respond by noting that no specific teaching have been pointed to by the Examiner. Instead, the Action merely posits that Appellants' specification "admits" that the starting materials were known. Appellants' specification merely teaches how one would obtain useful starting materials to prepare the present invention. It is submitted that Appellants' specification is the first to provide the requisite motivation to obtain these particular genetic starting materials and introduce them into corn. The Examiner's reliance upon Appellants' specification to provide the missing motivation constitutes impermissible hindsight reconstruction of the invention. *In re Deminski*, 230 U.S.P.Q. 313 (Fed. Cir. 1986).

***THE STANDARD OF OBVIOUSNESS UNDER  
O'FARRELL AND VAECK HAVE ALSO NOT BEEN MET***

Appellants would further request the Board to consider the case of *In re O'Farrell*, 7 U.S.P.Q.2d 1673, 1680 (Fed. Cir. 1988), which held that in order for a prior art teaching to obviate an invention, it must be shown that the teaching contains:

- (1) detailed enabling methodology for practicing the claimed invention;
- (2) a suggestion for modifying the prior art to practice the claimed invention; and
- (3) evidence suggesting that the invention would be successful.

In the more recent case of *In re Vaeck*, 20 U.S.P.Q. 1438 (Fed. Cir. 1991), the Federal Circuit took the *O'Farrell* doctrine a step further. In *Vaeck* the Federal Circuit stated that in order for an examiner to make out a *prima facie* case of obviousness two things must be shown: 1) that the prior art would have suggested to those of ordinary skill in the art that they should make the claimed composition; and 2) that the prior art must demonstrate a reasonable



expectation of success of the invention. Both the suggestion and the reasonable expectation of success must be founded in the prior art, not in the applicant's disclosure.

**1. There is No Evidence of Record that the Prior Art Would Predict that these Genetic Elements Would Function in Genetically Engineered Corn.**

The Examiner's argument appears to be based upon the assumption that the prior art teaches that the specified genes would function predictably in corn. While this standard overlooks the requirement for a motivation (discussed below), Appellants submit that there is no evidence of record that the prior art teaches the required predictability. The Action appears to suggest that the requirement for predictability is satisfied by Appellants' specification (Advisory Action, page 4, lines 75-79; "... previously known and isolated genes ... which have been shown to function transgenically, once inserted into maize ... are made obvious ..."). This would again constitute impermissible hindsight reconstruction. Absent an actual admission, nothing in appellants' specification can be considered pertinent to a prior art rejection. *In re Ruff*, 118 U.S.P.Q. 340 (CCPA 1958). The issue is whether the prior art teaches that the claimed genetic elements would function predictably when genetically engineered into corn – NOT whether the claimed elements do indeed function predictably in corn. No such teaching has been described by the Examiner.

**2. No Motivation For Preparing the Claimed Genetically Engineered Corn Species Has Been Shown**

The Examiner's reliance on the predictability issues overlooks the second part of the *O'Farrell* test; a motivation must be shown in the prior art for genetically engineering corn with the specific recited genetic elements.

In order to find the appealed claims obvious, the Examiner must demonstrate that the prior art provides a motivation for one of skill in the art to introduce the specified genes into corn. The claimed combination of elements in each of the claims -- which specify particular foreign genes introduced into the corn genome -- is a novel combination. As such, it is incumbent upon the Examiner to identify individual teaching(s) that would suggest the desirability of making each of the claimed combinations. As recently pointed out by the Federal Circuit, in assessing the patentability of a novel combination of otherwise old elements, "[t]he critical inquiry is whether there is something in the prior art as a whole to suggest the desirability, and thus the obviousness, of making the combination." *In re Newell*, 13 U.S.P.Q.2d 1248, 1250 (Fed. Cir. 1989).

The requirement that examiner's identify such a motivation is a longstanding patent law doctrine, even in biological cases. In the well-known case of *In re Bergel and Stock*, the CCPA stated it thusly:

The mere fact that it is possible to find two isolated disclosures that might be combined in such a way to produce a new compound does not necessarily render such production obvious unless the art also contains something to suggest the desirability of the proposed combination.

130 U.S.P.Q. 206 (CCPA 1961).

Motivation to combine elements can not be inferred from prior art that discloses one of the elements of the combination. This is the clear meaning of the relevant case law, such as *In re Gordon*, wherein the Federal Circuit stated:

The mere fact that the prior art could be so modified would not have made the modification obvious unless the prior art suggested the desirability of the modification.

221 U.S.P.Q. 1125 (Fed. Cir. 1984).

This doctrine has been routinely embraced by the PTO Board of Appeals. For example, in *Ex parte Kranz*, the Board made it clear that examiner's must identify either an explicit motivation, or a "compelling motivation based upon sound scientific principles." 19 U.S.P.Q.2d 1216, 1218 (BPAI 1990).

Here, the Examiner has taken the position that since the genes are known, and some have been placed into a particular plant species other than corn or previously expressed in a bacterium, that the introduction of the gene into corn in somehow inherently motivated. It is indeed Appellants' specification that provides for the first time the motivation for introducing the recited genes into corn -- and for this reason it is improper hindsight for the Examiner to rely upon Appellants' own specification to provide the necessary motivation. *Deminski*. This is certainly not the "compelling motivation" based upon "sound scientific principles" referred to by the *Kranz* Board, and certainly an insufficient basis for a *prima facie* case of obviousness.

**B. REJECTION OF CLAIMS AS OBVIOUS OVER THE '045 APPLICATION**

***SUMMARY OF THE REJECTION***

The Examiner next rejects the claims as obvious over the '045 specification. The basis for the obviousness rejection is essentially as set forth above with respect to the obviousness-type double patenting rejection, with the exception that the specification of the '045 application can be considered. Appellants respond initially by incorporating herein by reference the arguments set forth in the obviousness-type double patenting rejection.

Applicants would further address two specific points raised by the Examiner in the final rejection, and the Advisory Action.

***SECTION 103 AND DEUEL MANDATE THAT THE OBVIOUSNESS OF A METHOD OF MAKING AN INVENTION IS IRRELEVANT TO WHETHER THE INVENTION ITSELF IS OBVIOUS***

In the final Action (at page 6, lines 7-10), the Examiner states that since the various genes set forth in the claims are admittedly known in the art, there is nothing nonobvious about the claimed invention.

In response to this comment, Appellants reiterate that they are not claiming the various genes *per se*, but are claiming transgenic plants. In order to make out a *prima facie* rejection, the Examiner is bound to present prior art demonstrating that the claimed transgenic corn plants bearing the specified gene in their genome are obvious.

This is a case that falls under the ambit of the recent Federal Circuit case of *In re Deuel*, 34 U.S.P.Q.2d 1211 (Fed. Cir. 1995). The facts of *Deuel* are on all fours with those here: In *Deuel*, the applicant had cloned a particular gene encoding a known protein. The examiner

rejected the claim on the basis that the protein was known, and the methodology for cloning the gene from the known protein was routine in the art. The Federal Circuit reversed the examiner and the Board of Appeals, holding that the gene is not obvious simply because its encoded protein and the general cloning method were known. The court concluded that just because a method of preparing the cloned gene was known was irrelevant to the issue of obviousness. In the context of a claimed structure, it is imperative that the PTO demonstrate that the claimed structure is obvious, not that an obvious method for making the structure was known:

The PTO's focus on known methods for potentially isolating the claimed DNA molecules is also misplaced because the claims at issue define compounds, not methods. See *In re Bell*, 991 F.2d 781, 785, 26 U.S.P.Q.2d 1529, 1532 . . .

We today reaffirm the principle, stated in *Bell*, that the existence of a general method of isolating cDNA or DNA molecules is essentially irrelevant to the question whether the specific molecules themselves would have been obvious, in the absence of other prior art that suggests the claimed DNAs.

34 U.S.P.Q.2d at 1215.

The court also dismissed out of hand the concept argued by the PTO that since the underlying protein was known to have utility there existed a general motivation to clone the claimed gene:

The PTO's theory that one might have been motivated to try to do what Deuel in fact accomplished amounts to speculation and an impermissible hindsight reconstruction of the claimed invention . . . any motivation that existed [in the prior art] was a general one, to try to obtain a gene that was yet undefined and may have constituted many forms.

The court concluded that a general motivation to search for the gene of a known and characterized protein does not make *prima facie* obvious a gene that is subsequently obtained as a result of that search.

The legal issue decided in *Deuel* is not unlike the issue here: Whether the admittedly novel transgenic corn plants of the present invention are obvious over different transgenic corn plants of the '045 application, merely because the '045 application teaches techniques for obtaining transgenic corn plants in general. Both Section 103 and *Deuel* mandate a conclusion of nonobviousness. Section 103 states that "patentability shall not be negated by the manner in which the invention was made." The fact that a general technique exists for introducing a particular gene into corn is not relevant to the issue of obviousness under *Deuel*. *Deuel* and the earlier case of *Bell* reaffirm this principle, and each hold that reference must not be made to the obviousness of the method employed to make an invention. There is no teaching in the '045 specification as to the presently claimed plants, and admittedly no teaching in the '045 application with respect to the genes employed in the present invention.

In the Advisory Action, the Examiner appeared to agree with the foregoing proposition (Advisory Action, page 4, lines 74-75), yet attempts to draw a distinction by stating that the specific genes introduced by Appellants are otherwise known and have been "isolated and expressed." This is not a distinction, and is no different from the fact that in *Deuel* the protein encoded by the claimed gene was known and had been characterized. The *Deuel* court held that the protein was not relevant to the patentability of its corresponding gene, even though a technique was known for isolating the gene using the known characteristics of the protein. *Id.* Here, the fact that the previously known gene may have been isolated and expressed in a bacterium or even in another plant does not render obvious a transgenic corn plant that has been genetically engineered to express the gene. As

discussed above in responding to the obviousness-type double patenting rejection, it must be shown that a motivation existed in the prior art for making the claimed invention, and that it was predictable that the genetic element would function in a predictable manner in corn.

*O'Farrell and Vaeck.*

***THE FACT THAT STARTING MATERIALS WERE KNOWN IS NOT RELEVANT TO PATENTABILITY - THE ART DOES NOT TEACH OR SUGGEST THE PARTICULAR CLAIMED SPECIES***

The Advisory Action indirectly cites to teachings said to be contained in "references cited in [Appellants'] specification." (Advisory Action, page 4, lines 81-82). Appellants would again caution that while prior art references can be considered for what they teach, reliance upon Appellants' specification by the Examiner would constitute impermissible hindsight. Moreover, the Examiner has not listed in the rejection the specific references that are being relied upon.

If it is considered that the references and/or teachings incorporated into Appellants' specification are properly a part of this rejection, then it is submitted that in order to make out a *prima facie* rejection, the Examiner should consider each such teaching individually and it should be explained on the record by the Examiner how the reference teaches or suggests the subject matter of the claim. *Ex parte Goeddel*, 5 U.S.P.Q.2d 1449 (BPAI 1985). This has not been accomplished, and no *prima facie* case has been made.

Appellants will now review the subject matter of the individual claims that are to be considered separately for the purposes of the § 103 rejections, and further review the

references set forth in Appellants' specification and apparently relied upon by the Examiner to support this rejection.

Claim 47 -- Selectable or Screenable Marker Genes

Independent claim 47 is a Markush claim directed to a transgenic corn plant having one or more of two different selectable marker genes inserted in its genome, selected from:

an aequorin gene; and  
a gene encoding a cell wall protein (*e.g.*, an HPRG gene (claim 57)).

The Action fails to make out a *prima facie* case of obviousness with respect to any one of the foregoing transgenic corn species. No motivation for engineering these genetic elements into corn has been shown, and no expectation has been shown that they would function in a desired fashion even if successfully engineered into corn.

References relating to the foregoing starting materials are set forth in Exhibit D. Art exemplary of the aequorin gene starting material used in the preparation of the corn plants of claim 47 is reported in Knight *et al.*, which discusses the introduction and expression of the aequorin gene in transgenic tobacco. The Steifel *et al.* reference studies the expression of the maize HRGP gene in non-transgenic plants, and reports the cloning of the gene in an *E. coli* host (using an EMBL3 vector).

It is submitted that these references fail to render the claimed species *prima facie* obvious, as none meet the requirements of *O'Farrell*, *Vaeck*, *Bell* and *Deuel*. In particular, none of these references teach or suggest the introduction of the specified genetic element into corn (no motivation), and/or none evidence that the gene once introduced into corn



would successfully function to provide a benefit or solve a problem once introduced into corn.

Claim 60 -- Negatively Selectable Marker Genes

Independent claim 60 is a Markush claim directed to a transgenic corn plant comprising one or more of the following different genes in order to provide a negatively selectable marker in the resultant genetically engineered corn:

a cytosine deaminase gene;  
a T-DNA gene 2;  
an antisense bar gene; and  
an antisense nptII gene.

It is again submitted that the Action fails to make out a *prima facie* case of obviousness with respect to any one of the foregoing transgenic corn species. No motivation for engineering these genetic elements into corn has been shown, and no expectation has been shown that they would function in a desired fashion even if successfully engineered into corn.

References relating to the foregoing starting materials are set forth in Exhibit E. These three references, Stougaard et al., Depicker et al. and Xiang et al., individually relate to the expression of three of the foregoing genes in tobacco. It is submitted that the introduction and expression of these genes--the cytosine deaminase gene, the T-DNA gene 2 and the anti-nptII gene--in corn is novel and nonobvious.

The introduction and expression of the sense version of *bar* gene in corn has been described in the art of record (see, e.g., Ref. C76, Gordon-Kamm et al., *The Plant Cell*).

However, it is submitted that the purposeful expression of the antisense *bar* gene is not described in the prior art.

It is submitted that these teachings fail to render the claimed species *prima facie* obvious, as none meet the requirements of *O'Farrell*, *Vaeck*, *Bell* and *Deuel*. In particular, none of these references teach or suggest the introduction of the specified genetic element into corn (no motivation), and/or none teach or suggest that the gene once introduced into corn would function to provide a benefit and/or solve a problem.

Claim 61 -- Inducible or Tissue Specific Promoter or Enhancers

Independent claim 61 is a Markush claim directed to a transgenic corn plant comprising one or more of following different genetic control elements in order to provide an inducible or tissue specific promoter or enhancer capability to the resultant genetically engineered corn:

- an  $\alpha$ -tubulin promoter;
- an ocs promoter;
- an ABA-inducible promoter; and
- a turgor-inducible promoter.

The Action fails to make out a *prima facie* case of obviousness with respect to any one of the foregoing transgenic corn species. No motivation for engineering these genetic elements into corn has been shown, and no expectation has been shown that they would function in a desired fashion even if successfully engineered into corn.

References related to the foregoing starting materials are set forth in Exhibit F.

Guerrero *et al.* studies the expression of turgor-inducible genes in pea shoots, through cloning of the relevant genes in *E. coli*. Montoliu *et al.* similarly studies the expression of  $\alpha$ -tubulin genes in nontransgenic corn, using clones of genes prepared in *E. coli*. Mundy *et al.* relates to the ABA-inducible promoter of rice. Kononowicz *et al.* relates to the ocs promoter and its use in transgenic tobacco plants.

It is submitted that these references fail to render the claimed species *prima facie* obvious, as none meet the requirements of *O'Farrell*, *Vaeck*, *Bell* and *Deuel*. In particular, none of these references teach or suggest the introduction of the specified genetic element into corn (no motivation), and/or none teach or suggest that the gene once introduced into corn would function to provide a benefit or solve a problem.

#### Claim 62 -- Herbicide Resistance Genes

Independent claim 62 is directed to a transgenic corn plant comprising the *bxn* gene in order to provide herbicide resistance characteristics to the resultant genetically engineered corn. The Action fails to make out a *prima facie* case of obviousness with respect to transgenic corn bearing the *bxn* gene. No motivation for engineering the *bxn* gene into corn has been shown, and no expectation has been shown that it would function in a desired fashion even if successfully engineered into corn.

The Stalker *et al.* reference of Exhibit G, is believed to be exemplary of the preparation of the starting material *bxn* gene. Stalker *et al.* relates to the cloning and expression of the soil bacterium *bxn* gene in transgenic tobacco plants. It is submitted that this reference fails to render the claimed transgenic corn species *prima facie* obvious, as none

meet the requirements of *O'Farrell, Vaeck, Bell and Deuel*. In particular, references such as *Stalker et al.* fail to teach or suggest the introduction of the *bxn* gene into corn (no motivation), and such fail to evidence that the gene once introduced into corn would function to provide a benefit or solve a problem.

#### Claim 63 -- Insect Resistance Genes

Independent claim 63 is a Markush claim directed to a transgenic corn plant having one or more of following different genes in order to provide insect resistance characteristics to the resultant genetically engineered corn:

- an oryzacystatin gene;
- a wheat or barley amylase inhibitor gene;
- a lipoxygenase gene;
- an ecdysteroid UDP-glucosyl transferase gene; and
- a DIMBOA synthetic gene of the *bx* locus.

The Action again fails to make out a *prima facie* case of obviousness with respect to any one of the foregoing transgenic corn species. No motivation for engineering these genetic elements into corn has been shown, and no expectation has been shown that they would function in a desired fashion even if successfully engineered into corn.

References related to the foregoing starting materials are set forth in Exhibit H. Abe et al. concerns the cloning of an oryzacystatin gene in *E. coli*. Mundy et al. similarly relates to the cloning of the  $\alpha$ -amylase gene in *E. coli*. Yenofsky et al. also relates to cloning in *E. coli*, here the cloning of the lipoxygenase gene. O'Reilly et al. concerns the UDP-glycosyl transferase gene of an insect baculovirus. Dunn et al. recognizes the existence of the DIMBOA gene in non-transgenic corn.

It is submitted that these references fail to render the claimed species *prima facie* obvious, as none meet the requirements of *O'Farrell*, *Vaeck*, *Bell* and *Deuel*. In particular, none of these references teach or suggest the introduction of the specified genetic element into corn (no motivation), and none teach or suggest that the gene once introduced into corn would function to provide a benefit or solve a problem.

#### Claim 64 -- Disease Resistance Genes

Independent claim 64 is directed to a transgenic corn plant having a pathogenesis related (PR) protein gene in order to provide disease resistance to the resultant genetically engineered corn.

The Action again fails to make out a *prima facie* case of obviousness with respect to the foregoing transgenic corn species. No motivation for engineering a PR protein gene into corn has been shown, and no expectation has been shown that it would function in a desired fashion even if successfully engineered into corn. Attached as Exhibit I is the article of Bol et al., which reports the existence of PR proteins, but does not suggest the preparation of transgenic corn bearing such genes. The Bol et al. reference fails to render the claimed species *prima facie* obvious, as it fails to meet the requirements of *O'Farrell*, *Vaeck*, *Bell* and *Deuel*. In particular, it fails to teach the introduction of the specified genetic element into corn (no motivation), and fails to teach or suggest that the gene once introduced into corn would function to provide a benefit or solve a problem.

Claim 65 -- Stress Resistance Genes

Independent claim 65 is a Markush claim directed to a transgenic corn plant having one or more of following different genes in order to provide stress resistance characteristics to the resultant genetically engineered corn:

a glycerol-3-phosphate acetyltransferase gene;  
a superoxide dismutase gene; and  
a glutathione reductase gene.

The Action again fails to make out a *prima facie* case of obviousness with respect to any one of the foregoing transgenic corn species. No motivation for engineering these genetic elements into corn has been shown, and no expectation has been shown that they would function in a desired fashion even if successfully engineered into corn.

Attached as Exhibit J are various references concerning the foregoing genes, including Gupta et al., Wolter et al., Smith et al., Malan et al. and Bowler et al. It is submitted that these references merely concern the existence of these genes, and none teach or suggest transgenic corn bearing the specified genes. As such, it is submitted that these references fail to render the claimed species *prima facie* obvious, as none meet the requirements of *O'Farrell*, *Vaeck*, *Bell* and *Deuel*. In particular, none of these references teach or suggest the introduction of the specified genetic element into corn (no motivation), and none teach or suggest that the gene once introduced into corn would function to provide a benefit or solve a problem.

### Claim 66 -- Drought Resistance Genes

Independent claim 66 is a Markush claim directed to a transgenic corn plant having one or more of following different genes in order to provide drought resistance characteristics to the resultant genetically engineered corn:

- a mannitol-1-phosphate dehydrogenase gene;
- a trehalose-6-phosphate synthase gene;
- a myoinositol O-methyltransferase gene; and
- a Late Embryogenic Protein (LEA) gene.

The Action again fails to make out a *prima facie* case of obviousness with respect to any one of the foregoing transgenic corn species. No motivation for engineering these genetic elements into corn has been shown, and no expectation has been shown that they would function in a desired fashion even if successfully engineered into corn.

Attached as Exhibit K are various references concerning the foregoing genes, including Tarczynski et al., Kaasen et al., Vernon et al., and Dure et al. It is submitted that these references merely concern the existence of these genes, and none teach or suggest transgenic corn bearing the specified genes. As such, it is submitted that these references fail to render the claimed species *prima facie* obvious, as none meet the requirements of *O'Farrell, Vaeck, Bell and Deuel*. In particular, none of these references teach or suggest the introduction of the specified genetic element into corn (no motivation), and none teach or suggest that the gene once introduced into corn would function to provide a benefit or solve a problem.

### Claim 67 -- Grain Composition Genes

Independent claim 67 is a Markush claim directed to a transgenic corn plant having one or more of following different genes in order to provide certain grain composition characteristics to the resultant genetically engineered corn:

- an acetyl-CoA carboxylase gene;
- an ACP-acyltransferase gene;
- a b-ketoacyl-ACP synthase gene;
- an acyl carrier protein gene;
- a fatty acid desaturase gene;
- a fatty acid epoxidase gene;
- a fatty acid hydratase gene;
- a fatty acid dehydratase gene;
- a sense or antisense phytoene synthase gene;
- a sense or antisense phytoene desaturase gene;
- a sense or antisense lycopene synthase gene;
- a phytase gene;
- an ADP-glucose pyrophosphorylase gene;
- a starch synthase gene;
- a starch branching enzyme gene; and
- a sucrose synthase gene.

The Action again fails to make out a *prima facie* case of obviousness with respect to any one of the foregoing transgenic corn species. No motivation for engineering these genetic elements into corn has been shown, and no expectation has been shown that they would function in a desired fashion even if successfully engineered into corn. Attached as Exhibit L are references concerning various of the foregoing genes. It is submitted that these references merely concern the existence of these genes, and none teach or suggest transgenic corn bearing the specified genes. As such, it is submitted that these references fail to render the claimed species *prima facie* obvious, as none meet the requirements of *O'Farrell*, *Vaeck*, *Bell* and *Deuel*. In particular, none of these references teach or suggest the introduction of



the specified genetic element into corn (no motivation), and none teach or suggest that the gene once introduced into corn would function to provide a benefit or solve a problem.

***THE EXAMINER AGREES THAT TRANSGENIC CORN PLANTS BEARING GENES NOT PREVIOUSLY EXPRESSED IN PLANTS ARE NON-OBVIOUS, THUS MOOTING THE OBVIOUSNESS ISSUE WITH RESPECT TO CLAIMS DIRECTED TO THESE SPECIES***

In the Advisory Action, the Examiner appears to agree that claims directed to transgenic corn plants bearing genes NOT previously expressed in plants are patentable over the art. For example, at page 4, lines 86-89, it stated that "Applicants are correct that a gene which has no known function in a plant and which has not been demonstrated in the art to affect a plants phenotype, such as insect resistance or increase in storage proteins, would not provide—by its mere existence—motivation to place it in a plant ..."

Appellants agree with the foregoing proposition, but not with its implication. It is Appellants' position that the mere fact that a gene has been expressed in a plant other than corn, is not sufficient to render obvious a claim to transgenic corn expressing the specified gene for the reasons discussed above. Thus, Applicants respectfully disagree with the Examiner's implication that nonobviousness can only lie in the introduction of those gene species not previously expressed in another plant.

In light of the foregoing, it is respectfully submitted that the Examiner has not made out a prima facie rejection of the claims on the basis of obviousness over the '045 specification.

## C. REJECTION OF CLAIMS OVER THE GOLDMAN '073 PATENT

### *SUMMARY OF REJECTION*

Lastly, all of the claims have been rejected over the Goldman '073 patent in view of what the Examiner states is "Applicants acknowledged state of the prior art," presumably referring to the various references set forth in Appellants' specification that teach genetic starting materials employed in the preparation of the claimed transgenic corn plants. The Action takes the position that Goldman is enabling for the preparation of fertile, transgenic corn, and that Appellants' specification admits that the genes employed in the preparation of the claimed plants are known.

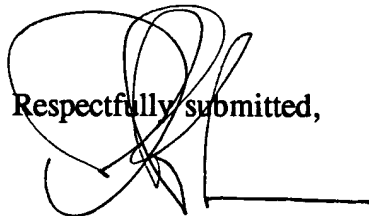
Appellants would first advise this Board that the appeal in related case USSN 07/565,844 (the '844 case), referred to in section IV. above, involves the issue of whether the '073 patent is enabling for the preparation of transgenic corn. In that related appeal, many arguments are presented to demonstrate that the '073 patent is not enabling for the preparation of fertile, transgenic corn. Appellants must reiterate here its position that Goldman is in no way enabling for the preparation of fertile, transgenic corn. However, Appellants do not contest in the present appeal the fact that general techniques were known for producing transgenic corn plants prior to the making of the present invention. Since corn transformation techniques were known at the time the present invention was made, the issue of whether the '073 patent is enabling is not ultimately relevant to the patentability of the present invention.) To insure against any acquiescence, Appellants incorporate by reference the arguments set forth in the '844 appeal, and have attached copies of relevant pages of that appeal brief as Exhibit M hereto.

In addition to the foregoing, to avoid duplication Appellants apply here the arguments set forth above in connection with the rejections over the '045 application. It is pointed out that the Goldman patent admittedly fails to in any way teach or suggest the particular transgenic corn species employed in the present invention, and thus is no more relevant in this regard than is the '045 specification addressed above. Because of the significant questions regarding the operability of Goldman, it is in fact a much less relevant reference.

**IX. SUMMARY AND CONCLUSION**

In light of the foregoing comments, appellants submit that the appealed claims meet the requirements for patentability. Therefore, appellants respectfully request that the Board reverse each of the rejections.

Respectfully submitted,



David L. Parker  
Reg. No. 32,165

Attorney for Appellants

ARNOLD, WHITE & DURKEE  
P.O. Box 4433  
Houston, Texas 77210  
(512) 418-3000

Date: October 18, 1995

## CLAIMS ON APPEAL

2. Cells obtained from the plant of any one of claims 47 or 60-67, wherein said cells comprise the introduced gene.
3. Progeny of the plant of any one of claims 47 or 60-67, wherein said progeny comprise the introduced gene.
4. Seeds obtained from the plant of claim 3, wherein said seeds comprise the introduced gene.
47. A fertile, transgenic maize plant, the genome of which has been augmented by the introduction of a DNA composition comprising a selectable or screenable marker gene selected from the group of genes consisting of an aequorin gene and a gene encoding a cell wall protein, so that the transgenic plant exhibits one or more phenotypic characteristics that render it identifiable over the corresponding untransformed maize plant which does not comprise said gene, and wherein said gene is transmittable through normal sexual reproduction of the transgenic maize plant to subsequent generation plants.
50. The transgenic maize plant of claim 48, wherein the autonomous replication vector comprises a transposable element.
51. The transgenic maize plant of claim 47, wherein the gene is positioned under the control of a promoter region comprising multiple copies of the 16 bp *ocs* enhancer element.
52. The transgenic maize plant of claim 47, wherein the selectable or screenable marker gene comprises non-expressed DNA.
53. The transgenic maize plant of claim 47, wherein the selectable or screenable marker gene comprises a dalapon dehalogenase (*deh*) gene.
54. The transgenic maize plant of claim 47, wherein the selectable or screenable marker gene comprises an anthranilate synthase gene that confers resistance to 5 methyl tryptophan.
55. The transgenic maize plant of claim 47, wherein the selectable or screenable marker gene comprises an aequorin gene.
56. The transgenic maize plant of claim 47, wherein the selectable or screenable marker gene comprises a gene encoding a cell wall protein.
57. The transgenic maize plant of claim 56, wherein the selectable or screenable marker gene comprises a gene encoding an HPRG.

58. The transgenic maize plant of claim 47, wherein the coding sequence of the gene is modified to improve expression in maize.

60. A fertile, transgenic maize plant, the genome of which has been augmented by the introduction of a DNA composition comprising a negatively-selectable marker selected from the group of genes consisting of a cytosine deaminase gene; a T-DNA gene 2; an antisense *bar* gene; and an antisense *nptII* gene, so that the transgenic plant exhibits one or more characteristics that render it identifiable over the corresponding untransformed maize plant which does not comprise said marker, and wherein said marker is transmittable through normal sexual reproduction of the transgenic maize plant to subsequent generation plants.

61. A fertile, transgenic maize plant, the genome of which has been augmented by the introduction of a DNA composition comprising an exogenous gene encoding a selected trait, the gene positioned under the control of an inducible or tissue-specific promoter or enhancer comprising an  $\alpha$ -tubulin promoter, an *ocs* promoter, an ABA-inducible promoter, or a turgor-inducible promoter, so that the transgenic plant exhibits one or more phenotypic characteristics that render it identifiable over the corresponding untransformed maize plant which does not comprise said gene, and wherein said gene is transmittable through normal sexual reproduction of the transgenic maize plant to subsequent generation plants.

62. A fertile, transgenic maize plant, the genome of which has been augmented by the introduction of a DNA composition comprising a gene encoding a herbicide resistance trait comprising a *bxn* gene, so that the transgenic plant exhibits one or more phenotypic characteristics that render it identifiable over the corresponding untransformed maize plant which does not comprise said gene, and wherein said gene is transmittable through normal sexual reproduction of the transgenic maize plant to subsequent generation plants.

63. A fertile, transgenic maize plant, the genome of which has been augmented by the introduction of a DNA composition comprising a gene encoding an insect resistance trait selected from the group of genes consisting of an oryzacystatin gene; a wheat or barley amylase inhibitor gene; a lipoxygenase gene; an ecdysteroid UDP-glucosyl transferase gene; and a DIMBOA synthetic gene of the *bx* locus, so that the transgenic plant exhibits one or more phenotypic characteristics that render it identifiable over the corresponding untransformed maize plant which does not comprise said gene, and wherein said gene is transmittable through normal sexual reproduction of the transgenic maize plant to subsequent generation plants.

64. A fertile, transgenic maize plant, the genome of which has been augmented by the introduction of a DNA composition comprising a gene encoding a pathogenesis related (PR) protein gene, so that the transgenic plant exhibits one or more phenotypic characteristics that render it identifiable over the corresponding untransformed maize plant which does not comprise said gene, and wherein said gene is transmittable through normal sexual reproduction of the transgenic maize plant to subsequent generation plants.

65. A fertile, transgenic maize plant, the genome of which has been augmented by the introduction of a DNA composition comprising a gene encoding a stress resistance trait selected from the group of genes consisting of a glycerol-3-phosphate acetyltransferase gene; a superoxide dismutase gene; and a glutathione reductase gene, so that the transgenic plant exhibits one or more phenotypic characteristics that render it identifiable over the corresponding untransformed maize plant which does not comprise said gene, and wherein said gene is transmittable through normal sexual reproduction of the transgenic maize plant to subsequent generation plants.

66. A fertile, transgenic maize plant, the genome of which has been augmented by the introduction of a DNA composition comprising a gene encoding a drought resistance trait selected from the group of genes consisting of a mannitol-1-phosphate dehydrogenase gene; a trehalose-6-phosphate synthase gene; a myoinositol 0-methyltransferase gene; and a Late Embryogenic Protein (LEA) gene, so that the transgenic plant exhibits one or more phenotypic characteristics that render it identifiable over the corresponding untransformed maize plant which does not comprise said gene, and wherein said gene is transmittable through normal sexual reproduction of the transgenic maize plant to subsequent generation plants.

67. A fertile, transgenic maize plant, the genome of which has been augmented by the introduction of a DNA composition comprising a gene encoding a grain composition trait selected from the group of genes consisting of an acetyl-CoA carboxylase gene; an ACP-acyltransferase gene; a  $\beta$ -ketoacyl-ACP synthase gene; an acyl carrier protein gene; a fatty acid desaturase gene; a fatty acid epoxidase gene; a fatty acid hydratase gene; a fatty acid dehydratase gene; a sense or antisense phytoene synthase gene; a sense or antisense phytoene desaturase gene; a sense or antisense lycopene synthase gene; a phytase gene; an ADP-glucose pyrophosphorylase gene; a starch synthase gene; a starch branching enzyme gene; and a sucrose synthase gene, so that the transgenic plant exhibits one or more phenotypic characteristics that render it identifiable over the corresponding untransformed maize plant which does not comprise said gene, and wherein said gene is transmittable through normal sexual reproduction of the transgenic maize plant to subsequent generation plants.

USSN  
07/508,045

FERTILE TRANSGENIC CORN PLANTS

Cross-Reference to Related Application

This application is a continuation-in-part of U.S. patent application Serial No. 07/467,983, filed January 22, 1990.

Field of the Invention

This invention relates to fertile transgenic plants of the species Zea mays (oftentimes referred to herein as maize or corn). The invention further relates to producing transgenic plants via particle bombardment and subsequent selection techniques which have been found to produce fertile transgenic plants.

Background of the Invention

Genetic engineering of plants, which entails the isolation and manipulation of genetic material (usually in the form of DNA or RNA) and the subsequent introduction of that genetic material into a plant or plant cells, offers considerable promise to modern agriculture and plant breeding. Increased crop food values, higher yields, feed value, reduced production costs, pest resistance, stress tolerance, drought resistance, the production of pharmaceuticals, chemicals and biological molecules as well as other beneficial traits are all potentially achievable through genetic engineering techniques. Once a gene has been identified, cloned, and engineered, it is still necessary to introduce it into a plant of interest in such a manner that the resulting plant is both fertile and capable of passing the gene on to its progeny.

A variety of methods have been developed and are currently available for the transformation of various plants and plant cells with DNA. Generally, these plants have been dicotyledonous, and some success has been reported with certain of the monocotyledonous cereals.

However, some species have heretofore proven untransformable by any method. Thus, previous to this discovery, no technology had been developed which would permit the production of stably transformed Zea mays plants in which the transforming DNA is heritable thereof. This failure in the art is well documented in the literature and has been discussed in a number of recent reviews (Potrykus, 1989; Weising et al., 1988; Cocking et al., 1987).

European Patent Publications 270,356 (McCabe et al.) and 275,069 (Arntzen et al.) describe the introduction of DNA into maize pollen followed by pollination of maize ears and formation of seeds. The plants germinated from these seeds are alleged to contain the introduced DNA, but there is no suggestion that the introduced DNA was heritable, as has been accomplished in the present invention. Only if the DNA introduced into the corn is heritable can the corn be used in breeding programs as required for successful commercialization of transgenic corn.

Graves et al. (1986) claim Agrobacterium-mediated transformation of Zea mays seedlings. The evidence was based upon assays known to be unreliable.

Despite extensive efforts to produce fertile transformed corn plants which transmit the transforming DNA to progeny, there have been no reported successes. Many previous failures have been based upon gene transfer to maize protoplasts, oftentimes derived from callus, liquid suspension culture cells, or other maize cells using a variety of transformation techniques. Although several of the techniques have resulted in successful transformation of corn cells, the resulting cells either could not be regenerated into corn plants or the corn plants produced were sterile (Rhodes et al. 1988) or, in some cases, it even turned out that the plants were, in fact, not transformed. Thus, while maize protoplasts and some other cells



have previously been transformed, the resulting transformants could not be regenerated into fertile transgenic plants.

On the other hand, it has been known that at least  
5 certain corn callus can be regenerated to form mature plants in a rather straightforward fashion and that the resulting plants are often fertile. However, no stable transformation of maize callus was ever achieved, i.e., there were no techniques developed which would permit a  
10 successful stable transformation of a regenerable callus. An example of a maize callus transformation technique which has been tried is the use of Agrobacterium-mediated transfer.

The art was thus faced with a dilemma. While it  
15 was known that corn protoplast and suspension culture cells could be transformed, no techniques were available which would regenerate the transformed protoplast into a fertile plant. While it was known that corn callus could be regenerated into a fertile plant, there were no techniques known  
20 which could transform the callus, particularly while not destroying the ability of the callus both to regenerate and to form fertile plants.

Recently, a new transformation technique has been created based upon the bombardment of intact cells and  
25 tissues with DNA-coated microprojectiles. The technique, disclosed in Sanford et al. (1987) as well as in EPO Patent Publication 331,855 of J. C. Sanford et al. based upon U.S. Serial No. 07/161,807, filed February 29, 1988, has been shown effective at producing transient gene expression in  
30 some plant cells and tissues including those from onion, maize (Klein et al. 1988a), tobacco, rice, wheat, and soybean, and stable expression has been obtained in tobacco and soybeans. In fact, stable expression has been obtained by bombardment of suspension cultures of Zea mays Black  
35 Mexican Sweet (Klein et al. 1989) which cultures are,

however, non-regenerable suspension culture cells, not the callus culture cells used in the process of the present invention.

No protocols have been published describing the introduction of DNA by a bombardment technique into cultures of regenerable maize cells of any type. No stable expression of a gene has been reported by means of bombardment of corn callus followed by regeneration of fertile plants and no regenerable fertile corn has resulted from DNA-coated microprojectile bombardment of the suspension cultures. Thus, the art has failed to produce fertile transformed corn plants heretofore.

A further stumbling block to the successful production of fertile transgenic maize plants has been in selecting those few transformants in such a manner that neither the regeneration capacity nor the fertility of the regenerated transformant are destroyed. Due to the generally low level of transformants produced by a transformation technique, the need for selection of the transformants is self-evident. However, selection generally entails the use of some toxic agent, e.g., herbicide or antibiotic, which may be detrimental to either the regenerability or the resultant plant fertility.

It is thus an object of the present invention to produce fertile, stably transgenic, Zea mays plants and seeds which transmit the introduced gene to progeny. It is a further object to produce such stably transgenic plants and seeds by a particle bombardment and a selection process which results in a high level of viability for at least a few transformed cells. It is a further object to produce fertile stably transgenic plants of other graminaceous cereals besides maize.

REFERENCES CITED

The references listed below are incorporated by reference herein.

- Armstrong, CL, et al. (1985) *Planta* 164:207-214
- 5 Callis, J, et al. (1987) *Genes & Develop* 1:1183-1200
- Chilton & Barnes (1983) *Nuc Acids Res* 11:364-385
- Chu, CC, et al. (1975) *Sci Sin (Peking)* 18:659-668
- Cocking, F, et al. (1987) *Science* 236:1259-1262
- DeWet et al. (1985) *Proc Natl Sci USA* 82:7870-7873
- 10 Freeling, JC, et al. (1976) *Maydica* XXI:97-112
- Graves, A, et al. (1986) *Plant Mol Biol* 7:43-50
- Green, C, et al. (1975) *Crop Sci* 15:417-421
- Green, C, et al. (1982) *Maize for Biological Research*,  
*Plant Mol Biol Assoc*, pp 367-372
- 15 Gritz, L, et al. (1983) *Gene* 25:179-188
- Guilley, H, et al. (1982) *Cell* 30:763-773
- Hallauer, AR, et al. (1988) *Corn and Corn Improvement*,  
 3rd ed., *Agronomy Society of America*, pp 469-564
- Jefferson, R, et al. (1987) *EMBO J* 6:3901-3907
- 20 Kamo, K, et al. (1985) *Bot Gaz* 146:327-334
- Klein, T, et al. (1989) *Plant Physiol* 91:440-444
- Klein, T, et al. (1988a) *Proc Natl Acad Sci USA*  
85:4305-9
- Klein, T, et al. (1988b) *Bio/Technology* 6:559-563
- 25 Lu, C, et al. (1982) *Theor Appl Genet* 62:109-112
- McCabe, D, et al. (1988) *Bio/Technology* 6:923-926
- Murashige, T, et al. (1962) *Physiol Plant* 15:473-497
- Neuffer, M, (1982) *Maize for Biological Research*, *Plant*  
*Mol Biol Assoc*, pp 19-30
- 30 Phillips, R, et al. (1988) *Corn and Corn Improvement*,  
 3rd ed., *Agronomy Society of America*, pp 345-387
- Potrykus, I (1989) *Trends in Biotechnology* 7:269-273
- Rhodes, CA, et al. (1988) *Science* 240:204-7
- Sambrook, J, et al. (1989) *Molecular Cloning: A Labor-*  
 35 *atory Manual*, 2nd ed., *Cold Spring Harbor Laboratory Press*

Sanford, J, et al. (1987) J Part Sci & Techn 5:27-37

Weising, K, et al., (1988) Ann Rev of Genetics 22:421-478

Yanisch-Perron, L, et al. (1985) Gene 33:109-119

5

### Summary of the Invention

The present invention relates to fertile transgenic Zea mays plants containing heterologous DNA, preferably chromosomally integrated heterologous DNA, which is heritable by progeny thereof.

The invention further relates to all products derived from transgenic Zea mays plants, plant cells, plant parts, and seeds.

The invention further relates to transgenic Zea mays seeds stably containing heterologous DNA and progeny which have inherited the heterologous DNA. The invention further relates to the breeding of transgenic plants and the subsequent incorporation of heterologous DNA into any Zea mays plant or line.

The invention further relates to a process for producing fertile transgenic Zea mays plants containing heterologous DNA. The process is based upon microprojectile bombardment, selection, plant regeneration, and conventional backcrossing techniques.

The invention further relates to a process for producing fertile transformed plants of graminaceous plants other than Zea mays which have not been reliably transformed by traditional methods such as electroporation, Agrobacterium, injection, and previous ballistic techniques.

The invention further relates to regenerated fertile mature maize plants obtained from transformed embryogenic tissue, transgenic seeds produced therefrom, and R1 and subsequent generations.

In preferred embodiments, this invention produces the fertile transgenic plants by means of a DNA-coated microprojectile bombardment of clumps of friable embryogenic callus, followed by a controlled regimen for selection of the transformed callus lines.

#### Brief Description of the Drawings

Figure 1A shows a map of plasmid vector pHYGI1 utilized in Example I. Figure 1B shows the relevant part of pHYGI1 encompassing the HPT coding sequence and associated regulatory elements. The base pair numbers start from the 5' nucleotide in the recognition sequence for the indicated restriction enzymes, beginning with the EcoRI site at the 5' end of the CaMV 35S promoter.

Figure 2 shows a map of plasmid vector pBII221 utilized in Example I.

Figure 3 is a Southern blot of DNA isolated from the PH1 callus line and an untransformed control callus line.

Figure 4 is a Southern blot of leaf DNA isolated from Ro plants regenerated from PH1 and untransformed callus.

Figure 5 is a Southern blot of leaf DNA isolated from R1 progeny of PH1 Ro plants and untransformed Ro plants.

Figure 6 is a Southern blot of DNA isolated from the PH2 callus line and an untransformed control callus line.

#### Description of the Preferred Embodiments

The present invention is directed to the production of fertile transgenic plants and seeds of the species Zea mays and to the plants, plant tissues, and seeds

derived from such transgenic plants, as well as the subsequent progeny and products derived therefrom. The transgenic plants produced herein include all plants of this species, including field corn, popcorn, sweet corn, flint  
5 corn and dent corn.

"Transgenic" is used herein to include any cell, cell line, callus, tissue, plant part or plant, the genotype of which has been altered beneficially by the presence of heterologous DNA that was introduced into the genotype  
10 by a process of genetic engineering, or which was initially introduced into the genotype of a parent plant by such a process and is subsequently transferred to later generations by sexual or asexual cell crosses or cell divisions. As used herein, "genotype" refers to the sum total of  
15 genetic material within a cell, either chromosomally, or extrachromosomally borne. Therefore, the term "transgenic" as used herein does not encompass the alteration of the genotype of Zea mays by conventional plant breeding methods or by naturally occurring events such as random cross-  
20 fertilization or spontaneous mutation.

By "heritable" is meant that the DNA is capable of transmission through a complete sexual cycle of a plant, i.e., it is passed from one plant through its gametes to its progeny plants in the same manner as occurs in normal  
25 corn.

The transgenic plants of this invention may be produced by (i) establishing a regenerable cell culture, preferably a friable embryogenic callus from the plant to be transformed, (ii) transforming said cell culture by a  
30 microprojectile bombardment technique, (iii) controllably identifying or selecting transformed cells, and (iv) regenerating fertile transgenic plants from the transformed cells. Some of the plants of this invention may be produced from the transgenic seed produced from the fertile

transgenic plants using conventional crossbreeding techniques to develop transgenic elite lines and varieties, or commercial hybrid seed containing heterologous DNA.

5 I. Plant Lines and Tissue Cultures

The cells which have been found particularly useful to produce the fertile transgenic maize plants herein are those callus cells which are regenerable, both before and after undergoing a selection regimen as detailed  
10 further below. Generally, these cells will be derived from meristematic tissue which contain cells which have not yet terminally differentiated. Such tissue in graminaceous cereals in general and in maize, in particular, comprise tissues found in juvenile leaf basal regions, immature  
15 tassels, immature embryos, and coleoptilar nodes. Preferably, immature embryos are used. Methods of preparing and maintaining callus from such tissue and plant types are well known in the art and details on so doing are available in the literature, c.f. Phillips et al. (1988), the disclosure of which is hereby incorporated by reference.  
20

The specific callus used must be able to regenerate into a fertile plant. The specific regeneration capacity of particular callus is important to the success of the bombardment/selection process used herein because during  
25 and following selection, regeneration capacity may decrease significantly. It is therefore important to start with cultures that have as high a degree of regeneration capacity as possible. Callus which is more than about 3 months and up to about 36 months of age has been found to have a  
30 sufficiently high level of regenerability and thus is preferred. The regenerative capacity of a particular culture may be readily determined by transferring samples thereof to regeneration medium and monitoring the formation of shoots, roots, and plantlets. The relative number of  
35 plantlets arising per petri dish or per gram fresh weight

of tissue may be used as a rough quantitative estimate of regeneration capacity. Generally, a culture which will produce at least one plant per gram of callus tissue is preferred.

5           While maize callus cultures can be initiated from a number of different plant tissues, the cultures useful herein are preferably derived from immature maize embryos which are removed from the kernels of an ear when the embryos are about 1-3 mm in length. This length generally  
10 occurs about 9-14 days after pollination. Under aseptic conditions, the embryos are placed on conventional solid media with the embryo axis down (scutellum up). Callus tissue appears from the scutellum after several days to a few weeks. After the callus has grown sufficiently, the  
15 cell proliferations from the scutellum may be evaluated for friable consistency and the presence of well-defined embryos. By "friable consistency" it is meant that the tissue is easily dispersed without causing injury to the cells. Tissue with this morphology is then transferred to  
20 fresh media and subcultured on a routine basis about every two weeks.

          The callus initiation media is solid because callus cannot be readily initiated in liquid medium. In preferred embodiments, the initiation/maintenance media is  
25 typically based on the N6 salts of Chu et al. (1975) as described in Armstrong et al. (1985) or the MS salts of Murashige et al. (1962). The basal medium is supplemented with sucrose and 2,4-dichlorophenoxyacetic acid (2,4-D). Supplements such as L-proline and casein hydrolysate have  
30 been found to improve the frequency of initiation of callus cultures, morphology, and growth. The cultures are generally maintained in the dark, though low light levels may also be used. The level of synthetic hormone 2,4-D, necessary for maintenance and propagation, should be generally  
35 about 0.3 to 3.0 mg/l.



Although successful transformation and regeneration has been accomplished herein with friable embryogenic callus, this is not meant to imply that other transformable regenerable cells, tissue, or organs cannot be employed to produce the fertile transgenic plants of this invention. The only actual requirement for the cells which are transformed is that after transformation they must be capable of regeneration of a plant containing the heterologous DNA following the particular selection or screening procedure actually used.

## II. DNA Used for Transformation

As used herein, the term "heterologous DNA" refers to a DNA segment that has been derived or isolated from one genotype, preferably amplified and/or chemically altered, and later introduced into a Zea mays genotype that may be the same Zea mays genotype from which the DNA was first isolated or derived. "Heterologous DNA" also includes completely synthetic DNA, and DNA derived from introduced RNA. Generally, the heterologous DNA is not originally resident in the Zea mays genotype which is the recipient of the DNA, but it is within the scope of the invention to isolate a gene from a given Zea mays genotype, and to subsequently introduce multiple copies of the gene into the same genotype, e.g., to enhance production of an amino acid.

Therefore, "heterologous DNA" is used herein to include synthetic, semi-synthetic, or biologically derived DNA which is introduced into the Zea mays genotype, and retained by the transformed Zea mays genotype. The DNA includes but is not limited to, non-plant genes such as those from bacteria, yeasts, animals or viruses; modified genes, portions of genes, chimeric genes, as well as genes from the same or different Zea mays genotype.

The heterologous DNA used for transformation herein may be circular or linear, double-stranded or single-stranded. Generally, the DNA is in the form of a plasmid and contains coding regions of beneficial heterologous DNA with flanking regulatory sequences which promote the expression of the heterologous DNA present in the resultant corn plant. For example, the heterologous DNA may itself comprise or consist of a promoter that is active in Zea mays, or may utilize a promoter already present in the Zea mays genotype that is the transformation target.

The compositions of and method for constructing heterologous DNA which can transform certain plants is well known to those skilled in the art, and the same compositions and methods of construction may be utilized to produce the heterologous DNA useful herein. The specific composition of the DNA is not central to the present invention and the invention is not dependent upon the composition of the specific transforming DNA used. Weising et al. (1988), the subject matter of which is incorporated herein by reference, describes suitable DNA components, which include promoters, polyadenylation sequences, selectable marker genes, reporter genes, enhancers, introns, and the like, as well as provides suitable references for compositions therefrom. Sambrook et al. (1989) provides suitable methods of construction.

Generally, the heterologous DNA will be relatively small, i.e., less than about 30 Kb to minimize any susceptibility to physical, chemical, or enzymatic degradation which is known to increase as the size of the DNA increases.

Suitable heterologous DNA for use herein includes all DNA which provides for, or enhances, a beneficial feature of the resultant transgenic corn plant. The DNA may encode proteins or antisense RNA transcripts in order to promote increased food values, higher yields, pest

resistance, disease resistance, and the like. For example, the DNA can encode a bacterial dap A for increased lysine production; Bacillus thuringiensis (BT) t-endotoxin or protease inhibitor for insect resistance; bacterial ESPS  
5 synthase for resistance to glyphosate herbicide; and chitinase or glucan endo-1,3-B-glucosidase for fungicidal properties. Aside from DNA sequences that serve as transcription units or portions thereof, useful DNA may be untranscribed, serving a regulatory or a structural function.  
10 Also, the DNA may be introduced to act as a genetic tool to generate mutants and/or assist in the identification, genetic tagging, or isolation of segments of corn DNA. Additional examples may be found in Weising, supra.

The heterologous DNA to be introduced into the  
15 plant further will generally contain either a selectable marker or a reporter gene or both to facilitate identification and selection of transformed cells. Alternatively, the selectable marker may be carried on a separate piece of DNA and used in a co-transformation procedure. Both selectable  
20 markers and reporter genes may be flanked with appropriate regulatory sequences to enable expression in plants. Useful selectable markers are well known in the art and include, for example, antibiotic and herbicide resistance genes. Specific examples of such genes are disclosed in  
25 Weising et al., supra. A preferred selectable marker gene is the hygromycin B phosphotransferase (HPT) coding sequence, which may be derived from E. coli and which confers resistance to the antibiotic hygromycin B. Other selectable markers include aminoglycoside phosphotrans-  
30 ferase gene of transposon Tn5 (AphII) which encodes resistance to the antibiotics kanamycin, neomycin, and G418, as well as those genes which code for resistance or tolerance to glyphosate, 1,2-dichloropropionic acid methotrexate, imidazolinones, sulfonylureas, bromoxynil, phosphonothricin  
35 and the like. Those selectable marker genes which confer

herbicide resistance or tolerance are also of commercial utility in the resulting transformed plants.

Reporter genes which encode for easily assayable marker proteins are well known in the art. In general, a reporter gene is a gene which is not present in or expressed by the recipient organism or tissue and which encodes a protein whose expression is manifested by some easily detectable property, e.g., phenotypic change or enzymatic activity. Examples of such genes are provided in Weising et al., supra. Preferred genes include the chloramphenicol acetyl transferase gene from Tn9 of E. coli, the beta-glucuronidase gene of the uidA locus of E. coli, and the luciferase genes from firefly Photinus pyralis.

The regulatory sequences useful herein include any constitutive, inducible, tissue or organ specific, or developmental stage specific promoter which can be expressed in the particular plant cell. Suitable such promoters are disclosed in Weising et al., supra. The following is a partial representative list of promoters suitable for use herein: regulatory sequences from the T-DNA of Agrobacterium tumefaciens, including mannopine synthase, nopaline synthase, and octopine synthase; alcohol dehydrogenase promoter from corn; light inducible promoters such as, ribulose-biphosphate-carboxylase small subunit gene from a variety of species; and the major chlorophyll a/b binding protein gene promoter; 35S and 19S promoters of cauliflower mosaic virus; developmentally regulated promoters such as the waxy, zein, or bronze promoters from maize; as well as synthetic or other natural promoters which are either inducible or constitutive, including those promoters exhibiting organ-specific expression or expression at specific development stage(s) of the plant.

Other elements such as introns, enhancers, polyadenylation sequences and the like, may also be present on the DNA. Such elements may or may not be necessary for the

function of the DNA, although they can provide a better expression or functioning of the DNA by affecting transcription, stability of the mRNA, or the like. Such elements may be included in the DNA as desired to obtain the optimal performance of the transforming DNA in the plant. For example, the maize Adh1S first intron may be placed between the promoter and the coding sequence of a particular heterologous DNA. This intron, when included in a DNA construction, is known to generally increase expression of a protein in maize cells. (Callis et al. 1987) However, sufficient expression for a selectable marker to perform satisfactorily can often be obtained without an intron. (Klein et al. 1989) An example of an alternative suitable intron is the shrunk-1 first intron of Zea mays. These other elements must be compatible with the remainder of the DNA constructions.

To determine whether a particular combination of DNA and recipient plant cells are suitable for use herein, the DNA may include a reporter gene. An assay for expression of the reporter gene may then be performed at a suitable time after the DNA has been introduced into the recipient cells. A preferred such assay entails the use of the E. coli beta-glucuronidase (GUS) gene (Jefferson et al. 1987). In the case of the microprojectile bombardment transformation process of the present invention, a suitable time for conducting the assay is about 2-3 days after bombardment. The use of transient assays is particularly important when using DNA components which have not previously been demonstrated or confirmed as compatible with the desired recipient cells.

### III. DNA Delivery Process

The DNA can be introduced into the regenerable maize cell cultures, preferably into callus cultures via a particle bombardment process. A general description of a

suitable particle bombardment instrument is provided in Sanford et al. (1987), the disclosure of which is incorporated herein by reference. While protocols for the use of the instrument in the bombardment of maize non-regenerable suspension culture cells are described in Klein et al. (1988a, 1988b, and 1989), no protocols have been published for the bombardment of callus cultures or regenerable maize cells.

In a microprojectile bombardment process, also referred to as a biolistic process, the transport of the DNA into the callus is mediated by very small particles of a biologically inert material. When the inert particles are coated with DNA and accelerated to a suitable velocity, one or more of the particles is able to enter into one or more of the cells where the DNA is released from the particle and expressed within the cell. While some of the cells are fatally damaged by the bombardment process, some of the recipient cells do survive, stably retain the introduced DNA, and express it.

The particles, called microprojectiles, are generally of a high density material such as tungsten or gold. They are coated with the DNA of interest. The microprojectiles are then placed onto the surface of a macroprojectile which serves to transfer the motive force from a suitable energy source to the microprojectiles. After the macroprojectile and the microprojectiles are accelerated to the proper velocity, they contact a blocking device which prevents the macroprojectile from continuing its forward path but allows the DNA-coated microprojectiles to continue on and impact the recipient callus cells. Suitable such instruments may use a variety of motive forces such as gunpowder or shock waves from an electric arc discharge (Swain et al. 1988). An instrument in which gunpowder is

the motive force is currently preferred and such is described and further explained in Sanford et al. (1987), the disclosure of which is incorporated herein by reference.

A protocol for the use of the gunpowder instrument is provided in Klein et al. (1988a, b) and involves two major steps. First, tungsten microprojectiles are mixed with the DNA, calcium chloride, and spermidine free base in a specified order in an aqueous solution. The concentrations of the various components may be varied as taught. The preferred procedure entails exactly the procedure of Klein et al. (1988b) except for doubling the stated optimum DNA concentration. Secondly, the DNA-coated microprojectiles, macroprojectiles, and recipient cells are placed in position in the instrument and the motive force is applied to the macroprojectiles. Parts of this step which may be varied include the distance of the recipient cells from the end of the barrel as well as the vacuum in the sample chamber. The recipient tissue is positioned 5 cm below the stopping plate tray.

The callus cultures useful herein for generation of transgenic plants should generally be about midway between transfer periods, and thus, past any "lag" phase that might be associated with a transfer to a new media, but also before reaching any "stationary" phase associated with a long time on the same plate. The specific tissue subjected to the bombardment process is preferably taken about 7-10 days after subculture, though this is not believed critical. The tissue should generally be used in the form of pieces of about 30 to 80, preferably about 40 to 60, mg. The clumps are placed on a petri dish or other surface and arranged in essentially any manner, recognizing that (i) the space in the center of the dish will receive the heaviest concentration of metal-DNA particles and the tissue located there is likely to suffer damage during bombardment and, (ii) the number of particles reaching a

cell will decrease (probably exponentially) with increasing distance of the cell from the center of the blast so that cells far from the center of the dish are not likely to be bombarded and transformed. A mesh screen, preferably of metal, may be laid on the dish to prevent splashing or ejection of the tissue. The tissue may be bombarded one or more times with the DNA-coated metal particles.

#### IV. Selection Process

Once the calli have been bombarded with the DNA and the DNA has penetrated some of the cells, it is necessary to identify and select those cells which both contain the heterologous DNA and still retain sufficient regenerative capacity. There are two general approaches which have been found useful for accomplishing this. First, the transformed calli or plants regenerated therefrom can be screened for the presence of the heterologous DNA by various standard methods which could include assays for the expression of reporter genes or assessment of phenotypic effects of the heterologous DNA, if any. Alternatively, and preferably, when a selectable marker gene has been transmitted along with or as part of the heterologous DNA, those cells of the callus which have been transformed can be identified by the use of a selective agent to detect expression of the selectable marker gene.

Selection of the putative transformants is a critical part of the successful transformation process since selection conditions must be chosen so as to allow growth and accumulation of the transformed cells while simultaneously inhibiting the growth of the non-transformed cells. The situation is complicated by the fact that the vitality of individual cells in a population is often highly dependent on the vitality of neighboring cells. Also, the selection conditions must not be so severe that the plant regeneration capacity of the callus cells and the



fertility of the resulting plant are precluded. Thus, the effects of the selection agent on cell viability and morphology should be evaluated. This may be accomplished by experimentally producing a growth inhibition curve for the given selective agent and tissue being transformed beforehand. This will establish the concentration range which will inhibit growth.

When a selectable marker gene has been used, the callus clumps may be either allowed to recover from the bombardment on non-selective media, or preferably, directly transferred to media containing that agent.

Selection procedures involve exposure to a toxic agent and may employ sequential changes in the concentration of the agent and multiple rounds of selection. The particular concentrations and cycle lengths are likely to need to be varied for each particular agent. A currently preferred selection procedure entails using an initial selection round at a relatively low toxic agent concentration and then later round(s) at higher concentration(s). This allows the selective agent to exert its toxic effect slowly over a longer period of time. Preferably, the concentration of the agent is initially such that about a 5-40% level of growth inhibition will occur, as determined from a growth inhibition curve. The effect may be to allow the transformed cells to preferentially grow and divide while inhibiting untransformed cells, but not to the extent that growth of the transformed cells is prevented. Once the few individual transformed cells have grown sufficiently, the tissue may be shifted to media containing a higher concentration of the toxic agent to kill essentially all untransformed cells. The shift to the higher concentration also reduces the possibility of non-transformed cells habituating to the agent. The higher level is preferably in the range of about 30 to 100% growth inhibition. The length of the first selection cycle may be from about 1

to 4 weeks, preferably about 2 weeks. Later selection cycles may be from about 1 to about 12 weeks, preferably about 2 to about 10 weeks. Putative maize transformants can generally be identified as proliferating sectors of tissue among a background of non-proliferating cells. The callus may also be cultured on non-selective media at various times during the overall selection procedure.

Once a callus sector is identified as a putative transformant, transformation can be confirmed by phenotypic and/or genotypic analysis. If a selection agent is used, an example of phenotypic analysis is to measure the increase in fresh weight of the putative transformant as compared to a control on various levels of the selective agent. Other analyses that may be employed will depend on the function of the heterologous DNA. For example, if an enzyme or protein is encoded by the DNA, enzymatic or immunological assays specific for the particular enzyme or protein may be used. Other gene products may be assayed by using a suitable bioassay or chemical assay. Other such techniques are well known in the art and are not repeated here. The presence of the gene can also be confirmed by conventional procedures, i.e., Southern blot or polymerase chain reaction (PCR) or the like.

## V. Regeneration of Plants and Production of Seed

Cell lines which have been shown to be transformed must then be regenerated into plants and the fertility of the resultant plants determined. Transformed lines which test positive by genotypic and/or phenotypic analysis are then placed on a media which promotes tissue differentiation and plant regeneration. Regeneration may be carried out in accordance with standard procedures well known in the art. The procedures commonly entail reducing the level of auxin which discontinues proliferation of a callus and

promotes somatic embryo development or other tissue differentiation. One example of such a regeneration procedure is described in Green et al. (1982). The plants are grown to maturity in a growth room or greenhouse and appropriate sexual crosses and selfs are made as described by Neuffer (1982).

Regeneration, while important to the present invention, may be performed in any conventional manner. If a selectable marker has been transformed into the cells, the selection agent may be incorporated into the regeneration media to further confirm that the regenerated plantlets are transformed. Since regeneration techniques are well known and not critical to the present invention, any technique which accomplishes the regeneration and produces fertile plants may be used.

#### VI. Analysis of R1 Progeny

The plants regenerated from the transformed callus are referred to as the R0 generation or R0 plants. The seeds produced by various sexual crosses of the R0 generation plants are referred to as R1 progeny or the R1 generation. When R1 seeds are germinated, the resulting plants are also referred to as the R1 generation.

To confirm the successful transmission and inheritance of the heterologous DNA in the sexual crosses described above, the R1 generation should be analyzed to confirm the presence of the transforming DNA. The analysis may be performed in any of the manners such as were disclosed above for analyzing the bombarded callus for evidence of transformation, taking into account the fact that plants and plant parts are being used in place of the callus.

## VII. Establishment of the Heterologous DNA in Other Maize Varieties

Fertile, transgenic plants may then be used in a conventional maize breeding program in order to incorporate the introduced heterologous DNA into the desired lines or varieties. Conventional breeding programs employ a conversion process (backcrossing). Methods and references for convergent improvement of corn are given by Hallauer et al., (1988) incorporated herein by reference. Briefly, conversion is performed by crossing the initial transgenic fertile plant to normal elite inbred lines. The progeny from this cross will segregate such that some of the plants will carry the heterologous DNA whereas some will not. The plants that do carry the DNA are then crossed again to the normal plant resulting in progeny which segregate once more. This backcrossing process is repeated until the original normal parent has been converted to a line containing the heterologous DNA and also possessing all other important attributes originally found in the parent. Generally, this will require about 6-8 generations. A separate backcrossing program will be generally used for every elite line that is to be converted to a genetically engineered elite line.

Generally, the commercial value of the transformed corn produced herein will be greatest if the heterologous DNA can be incorporated into many different hybrid combinations. A farmer typically grows several varieties of hybrids based on differences in maturity, standability, and other agronomic traits. Also, the farmer must select a hybrid based upon his physical location since hybrids adapted to one part of the corn belt are generally not adapted to another part because of differences in such traits as maturity, disease, and insect resistance. As such, it is necessary to incorporate the heterologous DNA into a large number of parental lines so that many hybrid

combinations can be produced containing the desirable heterologous DNA.

Corn breeding and the techniques and skills required to transfer genes from one line or variety to another are well known to those skilled in the art. Thus, introducing heterologous DNA into other lines or varieties can be readily accomplished by these breeding procedures whether or not they generate the appropriate calli.

#### 10 VIII. Uses of Transgenic Plants

The transgenic plants produced herein are expected to be useful for a variety of commercial and research purposes. Transgenic plants can be created for use in traditional agriculture to possess traits beneficial to the grower (e.g., agronomic traits such as pest resistance or increased yield), beneficial to the consumer of the grain harvested from the plant (e.g., improved nutritive content in human food or animal feed), or beneficial to the food processor (e.g., improved processing traits). In such uses, the plants are generally grown for the use of their grain in human or animal foods. However, other parts of the plants, including stalks, husks, vegetative parts, and the like, may also have utility, including use as part of animal silage or for ornamental purposes (e.g., Indian corn). Often, chemical constituents (e.g., oils or starches) of corn and other crops are extracted for foods or industrial use and transgenic plants may be created which have enhanced or modified levels of such components. The plants may also be used for seed production for a variety of purposes.

Transgenic plants may also find use in the commercial manufacture of proteins or other molecules encoded by the heterologous DNA contained therein, where the molecule of interest is extracted or purified from plant parts, seeds, and the like. Cells or tissue from the plants may

also be cultured, grown in vitro, or fermented to manufacture such molecules, or for other purposes (e.g., for research).

5 The transgenic plants may also be used in commercial breeding programs, or may be crossed or bred to plants of related crop species. Improvements encoded by the heterologous DNA may be transferred, e.g., from corn cells to cell of other species, e.g., by protoplast fusion.

10 The transgenic plants may have many uses in research or breeding, including creation of new mutant plants through insertional mutagenesis, in order to identify beneficial mutants that might later be created by traditional mutation and selection. The methods of the invention may also be used to create plants having unique  
15 "signature sequences" or other marker sequences which can be used to identify proprietary lines or varieties.

The following non-limiting examples are illustrative of the present invention. They are presented to better explain the general procedures which were used to  
20 prepare the fertile Zea mays plants of this invention which stably express the heterologous DNA and which transmit that DNA to progeny. All parts and percents are by weight unless otherwise specified. It must be recognized that a specific transformation event is a function of the amount  
25 of material subjected to the transformation procedure. Thus, when individual situations arise in which the procedures described herein do not produce a transformed product, repetition of the procedures will be required.

30 Example I.

Fertile transgenic Zea mays plants which contain heterologous DNA which is heritable were prepared as follows:

I. Initiation and maintenance of maize cell cultures which retain plant regeneration capacity

Friable, embryogenic maize callus cultures were initiated from hybrid immature embryos produced by pollination of inbred line A188 plants (University of Minnesota, Crop Improvement Association) with pollen of inbred line B73 plants (Iowa State University). Ears were harvested when the embryos had reached a length of 1.5 to 2.0 mm. The whole ear was surface sterilized in 50% v/v commercial bleach (2.63% w/v sodium hypochlorite) for 20 min. at room temperature. The ears were then washed with sterile, distilled, deionized water. Immature embryos were aseptically isolated and placed on nutrient medium initiation/-maintenance media with the root/shoot axis exposed to the medium. Initiation/maintenance media (hereinafter referred to as "F medium") consisted of N6 basal media (Chu 1975) with 2% (w/v) sucrose, 1.5 mg per liter 2,4-dichlorophenoxyacetic acid (2,4-D), 6 mM proline, and 0.25% Gelrite (Kelco, Inc., San Diego). The pH was adjusted to 5.8 prior to autoclaving. Unless otherwise stated, all tissue culture manipulations were carried out under sterile conditions.

The immature embryos were incubated at 26°C in the dark. Cell proliferations from the scutellum of the immature embryos were evaluated for friable consistency and the presence of well-defined somatic embryos. Tissue with this morphology was transferred to fresh media 10 to 14 days after the initial plating of the immature embryos. The tissue was then subcultured on a routine basis every 14 to 21 days. Sixty to eighty milligram quantities of tissue were removed from pieces of tissue that had reached a size of approximately one gram and transferred to fresh media. Subculturing always involved careful visual monitoring to be sure that only tissue of the correct morphology was maintained. The presence of somatic embryos ensured that

the cultures would give rise to plants under the proper conditions. The cell culture named AB12 used in this example was such a culture and had been initiated about 1 year before bombardment.

5

## II. Plasmids - pCHN1-1, pHYGI1, pBII221, and pLUC-1

The plasmids pCHN1-1, pHYGI1, and pLUC-1 were constructed in the vector pBS+ (Stratagene, Inc., San Diego, CA), a 3.2 Kb circular plasmid, using standard recombinant DNA techniques. pCHN1-1 contains the hygromycin B phosphotransferase (HPT) coding sequence from E. coli (Gritz et al. 1983) flanked at the 3' end by the nopaline synthase (nos) polyadenylation sequence of Agrobacterium tumefaciens (Chilton and Barnes 1983). Expression is driven by the cauliflower mosaic virus (CaMV) 35S promoter (Guilley et al. 1982), located upstream from the hygromycin coding sequence. The plasmid pHYGI1 was constructed by inserting the 553 bp Bcl-BamHI fragment containing the maize AdhIS first intron (Callis et al. 1987) between the CaMV 35S promoter and the hygromycin coding sequence of pCHN1-1. A map of pHYGI1 is provided as Figure 1. A sample of pHYGI1 was deposited at the American Type Culture Collection, Rockville, MD, USA, on March 16, 1990, under the provisions of the Budapest Treaty, and assigned accession number 40774.

pBII221 contains the E. coli B-glucuronidase coding sequence flanked at the 5' end by the CaMV 35S promoter and at the 3' end by the nos polyadenylation sequence. The plasmid was constructed by inserting the maize AdhIS first intron between the 35S promoter and the coding sequence of pBII221 (Jefferson et al. 1987). A map of pBII221 is provided as Figure 2.

pLUC-1 contains the firefly luciferase coding sequence (DeWet et al. 1987) flanked at the 5' end by the



CaMV 35S promoter and at the 3' end by the nos polyadenylation sequence. This plasmid was used solely as negative control DNA.

Plasmids were introduced into the embryogenic  
5 callus culture AB12 by microprojectile bombardment.

### III. DNA delivery process

The embryogenic maize callus line AB12 was subcultured 7 to 12 days prior to microprojectile bombardment.

10 AB12 was prepared for bombardment as follows. Five clumps of callus, each approximately 50 mg in wet weight were arranged in a cross pattern in the center of a sterile 60 x 15 mm petri plate (Falcon 1007). Plates were stored in a closed container with moist paper towels, throughout the  
15 bombardment process. Twenty-six plates were prepared.

Plasmids were coated onto M-10 tungsten particles (Biolistics) exactly as described by Klein et al. (1988b) except that, (i) twice the recommended quantity of DNA was used, (ii) the DNA precipitation onto the particles was  
20 performed at 0°C, and (iii) the tubes containing the DNA-coated tungsten particles were stored on ice throughout the bombardment process.

All of the tubes contained 25 µl 50 mg/ml M-10 tungsten in water, 25 µl 2.5 M CaCl<sub>2</sub>, and 10 µl 100 mM  
25 spermidine free base along with a total of 5 µl 1 mg/ml total plasmid content. When two plasmids were used simultaneously, each was present in an amount of 2.5 µl. One tube contained only plasmid pBII221; two tubes contained both plasmids pHYGI1 and pBII221; two tubes contained both  
30 plasmids pCHN1-1 and pBII221; and one tube contained only plasmid pLUC-1.

All tubes were incubated on ice for 10 min., pelletized by centrifugation in an Eppendorf centrifuge at

room temperature for 5 seconds, and 25  $\mu$ l of the supernatant was discarded. The tubes were stored on ice throughout the bombardment process. Each preparation was used for no more than 5 bombardments.

- 5           Macroprojectiles and stopping plates were obtained from Biolistics, Inc. (Ithaca, NY). They were sterilized as described by the supplier. The microprojectile bombardment instrument was obtained from Biolistics, Inc.

10           The sample plate tray was positioned at the position 5 cm below the bottom of the stopping plate tray of the microprojectile instrument, with the stopping plate in the slot nearest to the barrel. Plates of callus tissue prepared as described above were centered on the sample plate tray and the petri dish lid removed. A 7 x 7 cm  
15 square rigid wire mesh with 3 x 3 mm mesh and made of galvanized steel was placed over the open dish in order to retain the tissue during the bombardment. Tungsten/DNA preparation were sonicated as described by Biolistics, Inc. and 2.5  $\mu$ l was pipetted onto the top of the macroprojectiles.  
20           The instrument was operated as described by the manufacturer. The bombardments which were performed are summarized on Table 1.

Table 1.

25	2 x pBII221 prep	To determine transient expression frequency
	10 x PHYGI1/pBII221	As a potential positive treatment for transformation
30	10 x pCHN1-1/pBII221	As a potential positive treatment for transformation
	4 x pLUC-1	Negative control treatment

35

---

The two plates of callus bombarded with pBII221 were transferred plate for plate to F medium (with no hygromycin) and the callus cultured at 26°C in the dark.

After 2 days, this callus was then transferred plate for plate into 35 x 10 mm petri plates (Falcon 1008) containing 2 ml of GUS assay buffer which consists of 1 mg/ml 5-bromo-4-chloro-3-indolyl-beta-D-glucuronide (Research Organics),  
5 100 mM sodium phosphate pH 7.0, 5 mM each of potassium ferricyanide and potassium ferrocyanide, 10 mM EDTA, and 0.06% Triton X-100. These were incubated at 37°C for 3 days after which the number of blue cells was counted giving 291 and 477 transient GUS expressing cells in the  
10 two plates, suggesting that the DNA delivery process had also occurred with the other bombarded plates. These plates were discarded after counting since the GUS assay is destructive.

15 IV. Selection process

Hygromycin B (Calbiochem) was incorporated into the medium by addition of the appropriate volume of filter sterilized 100 mg/ml hygromycin B in water when the media had cooled to 45°C prior to pouring plates.

20 Immediately after all samples had been bombarded, callus from all of the plates treated with pHYGI1/pBII221, pCHN1-1/pBII221 and three of the plates treated with pLUC-1 were transferred plate for plate onto F medium containing 15 mg/l hygromycin B, (five pieces of callus per plate).  
25 These are referred to as round 1 selection plates. Callus from the fourth plate treated with pLUC-1 was transferred to F medium without hygromycin. This tissue was subcultured every 2-3 weeks onto nonselective medium and is referred to as unselected control callus.

30 After two weeks of selection, tissue appeared essentially identical on both selective and nonselective media. All callus from eight plates from each of the pHYGI1/pBII221 and pCHN1-1/pBII221 treatments and two plates of the control callus on selective media were trans-  
35 ferred from round 1 selection plates to round 2 selection

plates that contained 60 mg/l hygromycin. The round 2 selection plates each contained ten 30 mg pieces of callus per plate, resulting in an expansion of the total number of plates.

- 5           The remaining tissue on selective media, two plates each of pHYGI1/pBII221 and pCHN1-1/pBII221 treated tissue and one of control callus, were placed in GUS assay buffer at 37°C to determine whether blue clusters of cells were observable at two weeks post-bombardment. After 6  
10 days in assay buffer, this tissue was scored for GUS expression. The results are summarized on Table 2.

Table 2.

	<u>Treatment</u>	<u>Replicate</u>	<u>Observations</u>
15	pLUC-1		No blue cells
	pHYGI1/pBII221	Plate 1	11 single cells 1 four-cell cluster
		Plate 2	5 single cells
20	pCHN1-1/pBII221	Plate 1	1 single cell 2 two-cell clusters
		Plate 2	5 single cells 1 two-cell cluster 2 clusters of 8-10 cells
25			

- After 21 days on the round 2 selection plates, all viable portions of the material were transferred to round 3  
30 selection plates containing 60 mg/l hygromycin. The round 2 selection plates, containing only tissue that was apparently dead, were reserved. Both round 2 and 3 selection plates were observed periodically for viable proliferating sectors.

- 35           After 35 days on round 3 selection plates, both the round 2 and round 3 sets of selection plates were checked for viable sectors of callus. Two such sectors were observed proliferating from a background of dead

tissue on plates treated with pHYGI1/pBII221. The first sector named 3AA was from the round 3 group of plates and the second sector named 6L was from the round 2 group of plates. Both lines were then transferred to F medium without hygromycin.

After 19 days on F medium without hygromycin, the line 3AA grew very little whereas the line 6L grew rapidly. Both were transferred again to F medium for 9 days. The lines 3AA and 6L were then transferred to F medium containing 15 mg/l hygromycin for 14 days. At this point, line 3AA was observed to be of very poor quality and slow growing. The line 6L, however, grew rapidly on F medium with 15 mg/l hygromycin; the line was then subcultured to F medium without hygromycin.

After 10 days on F medium, an inhibition study of the line 6L was initiated. Callus of 6L was transferred onto F medium containing 1, 10, 30, 100, and 250 mg/l hygromycin B. Five plates of callus were prepared for each concentration and each plate contained ten approximately 50 mg pieces of callus. One plate of unselected control tissue was prepared for each concentration of hygromycin.

It was found that the line 6L was capable of sustained growth over 9 subcultures on 0, 10, 30, 100, and 250 mg/l hygromycin. The name of the line 6L was changed at this time from 6L to PH1 (Positive Hygromycin transformant 1).

Additional sectors were recovered at various time points from the round 2 and 3 selection plates. None of these were able to grow in the presence of hygromycin for multiple rounds, i.e., two or three subcultures.

#### V. Confirmation of transformed callus

To show that the PH1 callus had acquired the hygromycin resistance gene, a Southern blot of PH1 callus was prepared as follows: DNA was isolated from PH1 and

unselected control calli by freezing 2 g of callus in liquid nitrogen and grinding it to a fine powder which was transferred to a 30 ml Oak Ridge tube containing 6 ml extraction buffer (7M urea, 250 mM NaCl, 50 mM Tris-HCl pH 8.0, 20 mM EDTA pH 8.0, 1% sarcosine). To this was added 7 ml of phenol:chloroform 1:1, the tubes shaken and incubated at 37°C 15 min. Samples were centrifuged at 8K for 10 min. at 4°C. The supernatant was pipetted through miracloth (Calbiochem 475855) into a disposable 15 ml tube (American Scientific Products, C3920-15A) containing 1 ml 4.4 M ammonium acetate, pH 5.2. Isopropanol, 6 ml was added, the tubes shaken, and the samples incubated at -20°C for 15 min. The DNA was pelleted in a Beckman TJ-6 centrifuge at the maximum speed for 5 min. at 4°C. The supernatant was discarded and the pellet was dissolved in 500 µl TE-10 (10 mM Tris-HCl pH 8.0, 10 mM EDTA pH 8.0) 15 min. at room temperature. The samples were transferred to a 1.5 ml Eppendorf tube and 100 µl 4.4 M ammonium acetate, pH 5.2 and 700 µl isopropanol were added. This was incubated at -20°C for 15 min. and the DNA pelleted 5 min. in an Eppendorf microcentrifuge (12,000 rpm). The pellet was washed with 70% ethanol, dried, and resuspended in TE-1 (10 mM Tris-HCl pH 8.0, 1 mM EDTA).

The isolated DNA (10 µg) was digested with BamHI (NEB) and electrophoresed in a 0.8% w/v agarose gel at 15 V for 16 hrs in TAE buffer (40 mM Tris-acetate, 1 mM EDTA). The DNA within the gel was then depurinated by soaking the gel twice in 0.25 HCl for 15 min., denatured and cleaved by soaking the gel twice in 0.5 M NaOH/1.0 M NaCl 15 min., and neutralized by soaking the gel twice in 0.5 M Tris pH 7.4/3 M NaCl 30 min. DNA was then blotted onto a Nytran membrane (Shleicher & Shuell) by capillary transfer overnight in 6X SSC (20X SSC, 3 M NaCl, 0.3 M sodium citrate pH 7.0). The

membrane was baked at 80°C for 2 hrs under vacuum. Prehybridization treatment of the membrane was done in 6X SSC, 10X Denhardt's solution, 1% SDS, 50 µg/ml denatured salmon sperm DNA using 0.25 ml prehybridization solution per cm<sup>2</sup> of membrane. Prehybridization was carried out at 42°C overnight.

A <sup>32</sup>P labelled probe was prepared by random primer labelling with an Oligo Labelling Kit (Pharmacia) as per the supplier's instructions with <sup>32</sup>P-dCTP (ICN Radiochemicals). The template DNA used was the 1055 bp BamHI fragment of pHYG11, which is the HPT coding sequence. The fragment was gel purified and cut again with PstI (NEB) before labelling.

The hybridization was performed in 50% formamide, 6X SSC, 1% SDS, 50 µg/ml denatured salmon sperm DNA (Sigma), 0.05% sodium pyrophosphate and all of the isopropanol precipitated heat denatured probe (10<sup>7</sup> CPM/50 ng template). The hybridization was carried out at 42°C overnight.

The membrane was washed twice in 50 ml 6X SSC, 0.1% SDS 5 min. at room temperature with shaking, then twice in 500 ml 6X SSC, 0.1% SDS 15 min. at room temperature, then twice in 500 ml 1X SSC, 1% SDS 30 min. at 42°C, and finally in 500 ml 0.1X SSC, 1% SDS 60 min. at 65°C. Membranes were exposed to Kodak X-OMAT AR film in an X-OMATIC cassette with intensifying screens. As shown in Figure 3, a band was observed for PH1 callus at the expected position of 1.05 Kb, indicating that the HPT coding sequence was present. No band was observed for control callus.

To demonstrate that the hygromycin gene is incorporated into high molecular weight DNA, DNA isolated from PH1 callus and control callus was treated with (i) no restriction enzyme, (ii) BamHI, as described previously, or

(iii) PstI, which cuts the plasmid pHYGI1 only once within the HPT coding sequence. Samples were blotted and probed with the HPT coding sequence as described previously.

Undigested PH1 DNA only showed hybridization to the probe at the position of uncut DNA, demonstrating that the hygromycin gene is incorporated into high molecular weight DNA. The expected 1.05 Kb band for PH1 DNA digested with BamHI was observed, as had been shown previously. For PH1 DNA digested with PstI, a 5.9 Kb band would be expected if the hygromycin gene was present on an intact pHYGI1 plasmid. Two or more bands of variable size (size dependent on the position flanking PstI sites within the host DNA) would be expected if the gene was incorporated into high molecular weight DNA. Three bands were observed with approximate molecular sizes of 12, 5.1, and 4.9 Kb. This result demonstrates incorporation of the hygromycin gene into high molecular weight DNA. The intensity of the 4.9 Kb band is approximately twice as great as the other two bands, suggesting either partial digestion or possibly a tandem repeat of the HPT gene. No hybridization was observed for DNA from control callus in any of the above treatments.

These results prove that the HPT coding sequence is not present in PH1 callus as intact pHYGI1 or as a small non-chromosomal plasmid. They are consistent with incorporation of the hygromycin gene into high molecular weight DNA.

#### VI. Plant regeneration and production of seed

PH1 callus was transferred directly from all of the concentrations of hygromycin used in the inhibition study to RM5 medium which consists of MS basal salts (Murashige et al. 1962) supplemented with thiamine·HCl 0.5 mg/l, 2,4-D 0.75 mg/l, sucrose 50 g/l, asparagine 150 mg/l, and Gelrite 2.5 g/l (Kelco Inc., San Diego).



After 14 days on RM5 medium, the majority of PH1 and negative control callus was transferred to R5 medium which is the same as RM5 medium, except that 2,4-D omitted. These were cultured in the dark for 7 days at 26°C and transferred to a light regime of 14 hrs light and 10 hrs dark for 14 days at 26°C. At this point, plantlets that had formed were transferred to one quart canning jars (Ball) containing 100 ml of R5 medium. Plants were transferred from jars to vermiculite for 7 or 8 days before transplanting them into soil and growing them to maturity. A total of 65 plants were produced from PH1 and a total of 30 plants were produced from control callus.

To demonstrate that the introduced DNA had been retained in the Ro tissue, a Southern blot was performed as previously described on BamHI digested leaf DNA from three randomly chosen Ro plants of PH1. As shown in Figure 4, a 1.05 Kb band was observed with all three plants indicating that the HPT coding sequence was present. No band was observed for DNA from a control plant.

Controlled pollinations of mature PH1 plants were conducted by standard techniques with inbred Zea mays lines A188, B73, and Oh43. Seed was harvested 45 days post-pollination and allowed to dry further 1-2 weeks. Seed set varied from 0 to 40 seeds per ear when PH1 was the female parent and 0 to 32 seeds per ear when PH1 was the male parent.

#### VII. Analysis of the R1 progeny

The presence of the hygromycin resistance trait was evaluated by a root elongation bioassay, an etiolated leaf bioassay, and by Southern blotting. Two ears each from regenerated PH1 and control plants were selected for analysis. The pollen donor was inbred line A188 for all ears.

#### A. Root elongation bioassay

Seed was sterilized in a 1:1 dilution of commercial bleach in water plusalconox 0.1% for 20 min. in 125 ml Erlenmyer flasks and rinsed 3 times in sterile water and  
5 imbibed overnight in sterile water containing 50 mg/ml captan by shaking at 150 rpm.

After imbibition, the solution was decanted from the flasks and the seed transferred to flow boxes (Flow Laboratories) containing 3 sheets of H<sub>2</sub>O saturated germination  
10 tion paper. A fourth sheet of water saturated germination paper was placed on top of the seed. Seed was allowed to germinate 4 days.

After the seed had germinated, approximately 1 cm of the primary root tip was excised from each seedling and  
15 plated on MS salts, 20 g/l sucrose, 50 mg/l hygromycin, 0.25% Gelrite, and incubated in the dark at 26°C for 4 days.

Roots were evaluated for the presence or absence of abundant root hairs and root branches. Roots were  
20 classified as transgenic (hygromycin resistant) if they had root hairs and root branches, and untransformed (hygromycin sensitive) if they had limited numbers of branches. The results are shown in Table 3, hereinbelow.

#### 25 B. Etiolated leaf bioassay

After the root tips were excised as described above, the seedlings of one PH1 ear and one control ear were transferred to moist vermiculite and grown in the dark for 5 days. At this point, 1 mm sections were cut from the  
30 tip of the coleoptile, surface sterilized 10 seconds, and plated on MS basal salts, 20 g/l sucrose, 2.5 g/l Gelrite with either 0 (control) or 100 mg/l hygromycin and incubated in the dark at 26°C for 18 hrs. Each plate contained duplicate sections of each shoot. They were then incubated  
35 in a light regimen of 14 hrs light 10 hrs dark at 26°C for

48 hrs, and rated on a scale of from 0 (all brown) to 6 (all green) for the percent of green color in the leaf tissue. Shoots were classified as untransformed (hygromycin sensitive) if they had a rating of zero and classified as transformed (hygromycin resistant) if they had a rating of 3 or greater. The results are shown in Table 1, hereinbelow.

### C. Southern blots

10               Seedlings from the bioassays were transplanted to soil and were grown to sexual maturity. DNA was isolated from 0.8 g of leaf tissue about 3 weeks after transplanting to soil and probed with the HPT coding sequence as described previously. Plants with a 1.05 Kb band present in  
15 the Southern blot were classified as transgenic. As shown in Figure 5, two out of seven progeny of PH1 plant 3 were transgenic as were three out of eight progeny of PH1 plant 10. The blot results correlated precisely with data from  
20 the bioassays, confirming that the heterologous DNA was transmitted through one complete sexual life cycle. All data are summarized in Table 3.

Table 3  
ANALYSIS OF PH1 R1 PLANTS

	<u>PH1</u> <u>PLANT</u>	<u>ROOT</u> <u>ASSAY</u>	<u>LEAF</u> <u>ASSAY</u>	<u>BLOT</u> <u>_____</u>	<u>CONT</u> <u>PLANT</u>	<u>ROOT</u> <u>ASSAY</u>	<u>LEAF</u> <u>ASSAY</u>	<u>BLOT</u> <u>_____</u>
5	3.1	+	ND	+	4.1	-	ND	ND
	3.2	-	ND	-	4.2	-	ND	ND
	3.3	-	ND	-	4.3	-	ND	ND
	3.4	-	ND	-	4.4	-	ND	ND
10	3.5	-	ND	-	4.5	-	ND	ND
	3.6	+	ND	+	4.6	-	ND	ND
	3.7	-	ND	-	4.7	-	ND	ND
					2.1	-	ND	-
	10.1	+	+	+	1.1	-	-	-
15	10.2	+	+	+	1.2	-	-	ND
	10.3	-	-	ND	1.3	-	-	ND
	10.4	-	-	-	1.4	-	-	ND
	10.5	-	-	-	1.5	-	-	ND
	10.6	-	-	-	1.6	-	-	ND
20	10.7	-	-	-	1.7	-	-	ND
	10.8	ND	+	+	1.8	-	-	ND

Key: + = transgenic; - = nontransgenic; ND = not done

25

#### Example II.

The procedure of Example I was repeated with minor modifications.

#### 30 I. Plant Lines and Tissue Cultures

The embryogenic maize callus line, AB12, was used as in Example I. The line had been initiated about 18 months before the actual bombardment occurred.

#### 35 II. Plasmids

The plasmids pBII221 and pHYGI1 described in Example I were used.

### III. DNA Delivery Process

Callus was bombarded exactly as in Example I except that the DNA used in the tungsten/DNA preparations differed. All of the tubes contained 25  $\mu$ l 50 mg/ml M-10 tungsten in water, 25  $\mu$ l 2.5 M  $\text{CaCl}_2$ , and 10  $\mu$ l 100 mM spermidine free base along with a total of 5  $\mu$ l 1 mg/ml total plasmid content. One tube contained only plasmid pBII221; two tubes contained only plasmid pHYGI1; and one tube contained no plasmid but 5  $\mu$ l TE-1 (10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0).

The following bombardments were done: 2 x pBII221 prep (for transient expression); 7 x pYGI1 prep (potential positive treatment); and 3 x TE prep (negative control treatment).

After all the bombardments were performed, the callus from the pBII221 treatment was transferred plate for plate to F medium as five 50 mg pieces. After 2 days, the callus was placed into GUS assay buffer as per Example I. Numbers of transiently expressing cells were counted and found to be 686 and 845 GUS positive cells, suggesting that the particle delivery process had occurred in the other bombarded plates.

### IV. Selection of Transformed Callus

After bombardment, the callus from the pYGI1 treatments was placed onto round 1 selection plates, F medium containing 15 mg/l hygromycin, as ten 25 mg pieces per plate (different from Example I). The same was done for two of the plates bombarded with the TE preparation (selected control callus). One plate of callus bombarded with the TE preparation was placed onto F medium with no hygromycin; this callus was maintained throughout the

ongoing experiment as a source of control tissue (unselected control callus).

After 13 days, the callus on round 1 selection plates was indistinguishable from unselected control  
5 callus. All of the callus was transferred from round 1 selection plates to round 2 selection plates containing 60 mg/l hygromycin. An approximate five-fold expansion of the numbers of plates occurred.

The callus on round 2 selection plates had  
10 increased substantially in weight after 23 days, but at this time appeared close to dead. All of the callus was transferred from round 2 selection plates to round 3 selection plates containing 60 mg/l hygromycin. This transfer  
15 of all material from round 2 to round 3 selection differs from Example I in which only viable sectors were transferred and the round 2 plates reserved.

At 58 days post-bombardment, three live sectors were observed proliferating from the surrounding dead tissue. All three lines were from PHYGI1 treatments and  
20 were designated 24C, 56A, and 55A.

After 15 days on maintenance medium, growth of the lines was observed. The line 24C grew well whereas lines 55A and 56A grew more slowly. All three lines were transferred to F medium containing 60 mg/l hygromycin. Unselected control callus from maintenance medium was plated to  
25 F medium having 60 mg/l hygromycin.

After 19 days on 60 mg/l hygromycin, the growth of line 24C appeared to be entirely uninhibited, with the control showing approximately 80% of the weight gain of  
30 24C. The line 56A was completely dead, and the line 55A was very close to dead. The lines 24C and 55A were transferred again to F 60 mg/l hygromycin as was the control tissue.

After 23 days on 60 mg/l hygromycin, the line 24C  
35 again appeared entirely uninhibited. The line 55A was

completely dead, as was the negative control callus on its second exposure to F medium having 60 mg/l hygromycin.

At 88 days post-bombardment, a sector was observed proliferating among the surrounding dead tissue on the round 3 selection plates. The callus was from a plate bombarded with pHYGI1 and was designated 13E. The callus was transferred to F medium and cultured for 19 days. Portions of the callus were then transferred to (i) F media containing 15 mg/l hygromycin, and (ii) F medium containing 60 mg/l hygromycin. Control callus was plated on F media with 15 mg/l hygromycin. After 14 days of culture, the callus line 13E appeared uninhibited on both levels of hygromycin. The control callus appeared to have about 80% of the weight gain of 13E. The callus lines were transferred to fresh media at the same respective levels of hygromycin.

#### V. Confirmation of Transformed Callus

A Southern blot was prepared from BamHI-digested DNA from the line 24C. As shown in Figure 6, a band was observed for the line 24C at the expected size of 1.05 Kb showing that the line 24C contained the HPT coding sequence. No band was observed for DNA from control tissue. The name of the callus line 24C was changed to PH2.

To demonstrate that the hygromycin gene is incorporated into high molecular weight DNA, DNA isolated from PH2 callus and control callus was treated with (i) no restriction enzyme, (ii) BamHI, as described previously, or, (iii) PstI, which cuts the plasmid pHYGI1 only once within the HPT coding sequence. Samples were blotted and probed with the HPT coding sequence as described previously.

Undigested PH2 DNA only showed hybridization to the probe at the position of uncut DNA, demonstrating that

the hygromycin gene is incorporated into high molecular weight DNA. The expected 1.05 Kb band for PH2 DNA digested with BamHI was observed, as had been shown previously. For PH2 DNA digested with PstI, a 5.9 Kb band would be expected if the hygromycin gene was present on an intact pHYGI1 plasmid. Two or more bands of variable size (size dependent on the position of flanking PstI sites within the host DNA) would be expected if the gene was incorporated into high molecular weight DNA. Two bands were observed with approximate molecular sizes of 6.0 and 3.0 Kb. This result is consistent with incorporation of the hygromycin gene into high molecular weight DNA. No hybridization was observed for DNA from control callus in any of the above treatments.

These results prove that the HPT coding sequence is not present in PH2 callus as intact pHYGI1 or as a small non-chromosomal plasmid. They are consistent with incorporation of the hygromycin gene into high molecular weight DNA.

20

#### VI. Plant Regeneration and Production of Seed

The line PH2, along with unselected control callus, were placed onto RM5 medium to regenerate plants as in Example I. After 16 days, the callus was transferred to R5 medium as in Example I. After 25 d on R5 medium, plantlets were transferred to R5 medium and grown up for 20 days. At this point, plantlets were transferred to vermiculite for one week and then transplanted into soil where they are being grown to sexual maturity.

30

#### Example III.

The procedure of Example II was repeated exactly except that different plasmids were used.

The plasmids pBII221 and pHYGI1 described in Example I were used as well as pMS533 which is a plasmid

35



that contains the insecticidal Bacillus thuringiensis endotoxin (BT) gene fused in frame with the neomycin phosphotransferase (NPTII) gene. At a position 5' from the fusion gene are located segments of DNA from the CaMV and  
 5 nopaline synthase promoters. At a position 3' from the fusion gene are segments of DNA derived from the tomato protease inhibitor I gene and the poly A region of the nopaline synthase gene.

Callus was bombarded exactly as in Example I  
 10 except that the DNA used in the tungsten/DNA preparations differed. Two tubes contained plasmids pHYGI1 and pMS533 and one tube contained plasmids pHYGI1 and pMS533 and one tube contained no plasmid but contained 5  $\mu$ l TE-1 (10 mM Tris-HCl pH 8.0, 1 mM EDTA pH 8.0).

15 The following bombardments were done: 9 x pYHYGI1/pMS533 (potential positive treatment) and 2 x TE prep (control treatment).

After bombardment, the callus from the pYHYGI1/-pMS533 treatments was placed onto round 1 selection plates,  
 20 F medium containing 15 mg/l hygromycin, as ten 25 mg pieces per plate. The same was done for one of the plates bombarded with the TE preparation (selected control callus). One plate of callus bombarded with the TE preparation was placed onto F medium with no hygromycin; this callus was  
 25 maintained throughout the ongoing experiment as a source of control tissue (unselected control callus).

After 12 days, the callus on round 1 selection plates appeared to show about 90% of the weight gain of the unselected control callus. All of the callus was trans-  
 30 ferred from round 1 selection plates to round 2 selection plates containing 60 mg/l hygromycin as ten 30 mg pieces per plate. After 22 days of selection on round 2 selection plates, the callus appeared completely uninhibited. All of the callus was transferred from round 2 selection plates to  
 35 round 3 selection plates containing 60 mg/l hygromycin.

At 74 days post-bombardment, a single viable sector was observed proliferating from the surrounding necrotic tissue. The callus line was from PHYGI1/pMS533 treated material and was designated 86R. The callus line  
5 86R was transferred to F medium.

After 24 days, the callus line 86R had grown substantially. Portions of the callus were then transferred to (i) F media containing 15 mg/l hygromycin, and (ii) F media containing 60 mg/l hygromycin. Control callus  
10 was plated on F media with 15 mg/l hygromycin.

After 19 days of culture, the callus line 86R appeared to grow rapidly and was uninhibited on both levels of hygromycin. The control callus appeared to have only about 50% of the weight gain of 86R. The callus lines were  
15 transferred to fresh media at the same respective levels of hygromycin to further test the resistance of the callus line 86R. After 26 days of culture, the callus line 86R appeared uninhibited on 60 mg/l hygromycin.

Southern blots were performed on DNA isolated from  
20 the callus line 86R and control callus to confirm the presence of the hygromycin resistance gene and to determine whether the BT gene was present.

For detection of the HPT coding sequence, DNA isolated from 86R callus and control callus was digested  
25 with the restriction enzymes BamHI, XhoI, or PstI as described in Examples I and II. After hybridization with a probe prepared from the HPT coding sequence, the following bands were observed. For the BamHI digest, bands were observed at the expected size of 1.05 Kb as well as at  
30 approximately 3.0 and 2.3 Kb. This result demonstrates that the HPT coding sequence is present in the callus line 86R. The additional bands at 3.0 and 2.3 Kb indicate that either digestion was incomplete or that multiple rearranged copies are present. For the XhoI digest, a single band was  
35 observed at approximately 5.1 Kb. Because XhoI does not

cut pHYGI1, this suggests incorporation of the hygromycin construct into DNA different than pHYGI1. For the PstI digestion, a large band was observed at approximately 5.1 Kb. This band appeared to be two fragments of similar molecular weight. Two or more bands would be expected from a PstI digestion if the gene was incorporated into high molecular weight DNA. In no case was hybridization observed for DNA from control callus for any of the above-mentioned digestions.

10 For detection of the BT gene, a Southern blot was carried out on DNA isolated from 86R and control callus digested with the enzymes BamHI and XhoI in combination. A BamHI, XhoI co-digestion liberates the 1.8 Kb BT coding sequence from the pMS533 construction used in this trans-formation. The blot prepared was hybridized to a probe prepared from the 1.8 Kb BT coding sequence. A band was observed for 86R DNA at the expected size of 1.8 Kb whereas no hybridization was observed for control DNA. Additional bands of much lesser intensity were also observed for 86R DNA. This result demonstrates that the BT coding sequence is present in the callus line 86R. This further demonstrates the introduction into maize of an unselected gene with potential commercial value. The name of callus line 86R was changed to CB1.

25 Plants are being regenerated from CB1 callus and control callus as described in Example I.

#### COMPARATIVE EXAMPLE A

30 The basic procedures of Examples I-III have been attempted except varying the selection regime or the form of the callus. These other attempts, which are detailed in Table 4 below, were not successful. Since they were not repeated several times, it is not known whether they can be made to work. In all of the procedures, no viable sectors were observed. In the Table, "sieved" indicates that the

35

callus was passed through an 860 micron sieve before bombardment. The selection agent was hygromycin for each case except when pMXTI1 was the plasmid and methotrexate the selection agent.

5

Table 4  
Summary of Comparative Example A

	<u>Recip.</u> <u>Tissue</u>	<u>Plasmids</u>	<u>Recov.</u> <u>Period</u>	<u>Round 1</u> <u>Level</u>	<u>Round 1</u> <u>Period</u>	<u>Round 2</u> <u>Level</u>	<u>Round 2</u> <u>Period</u>
10	Clumps	pCHN1-1 pBII221	13	60	21	60	81
	Clumps	pCHN1-1 pBII221	14	100	22	-	-
15	Clumps	pHYGI1 pBII221	8	60	19	30	132
	Clumps	pCHN1-1 pBII221	0	30	22	60	109
	Clumps	pMTXI1 pBII221	8	3	103	-	-
20	Sieved	pCHN1-1 pBII221	13	-	-	-	-

The invention has been described with reference to various specific and preferred embodiments and techniques. However, it should be understood that many variations and modifications may be made while remaining within the spirit and scope of the invention.

25

**WHAT IS CLAIMED IS:**

1. A fertile transgenic Zea mays plant containing heterologous DNA which is heritable.
2. The plant of claim 1 wherein the heterologous DNA is chromosomally integrated.
3. The plant of claim 1 wherein the heterologous DNA is expressed.
4. The plant of claim 1 wherein the heterologous DNA encodes a protein.
5. The plant of claim 1 wherein the heterologous DNA comprises a promoter.
6. The plant of claim 1 selected from the group consisting essentially of field corn, popcorn, sweet corn, flint corn, and dent corn.
7. The plant of claim 1 wherein said heterologous DNA comprises a DNA sequence selected from the group consisting of a bacterial dap A gene, a BT-endotoxin gene, a protease inhibitor gene, a bacterial ESPS synthase gene, a chitinase gene, a glucan endo-1,3-B-glucosidase gene, structural DNA, regulatory sequence, an identification sequence and a sequence encoding antisense RNA.
8. The plant of claim 1 wherein said heterologous DNA encodes a beneficial trait to the plant.
9. The plant of claim 8 wherein said beneficial trait is selected from the group consisting of promoting

increased food value, higher yield, reduced production cost, pest resistance, stress tolerance, drought resistance, and disease resistance.

10. The plant of claim 1 which expresses a selectable marker gene or a reporter gene.
11. The plant of claim 10 wherein the selectable marker gene confers resistance or tolerance to a compound selected from the group consisting of hygromycin, kanamycin, G418, 1,2-dichloropropionic acid, neomycin, phosphonotricin, glyphosate, methotrexate, imidazolinone, chlorsulfuron, and bromoxynil.
12. The plant of claim 11 wherein the selectable marker gene confers resistance or tolerance to hygromycin.
13. The plant of claim 10 which expresses a reporter gene.
14. The seed produced by the plant of claim 1 which has inherited the heterologous DNA.
15. The grain produced from the plant of claim 1 wherein the heterologous DNA increases the feed value or the food value of said grain.
16. The R1 and subsequent generations derived from the plant of claim 1.
17. A process for producing a fertile transgenic Zea mays plant which stably expresses heterologous DNA which is heritable, wherein the process comprises the steps of (i) establishing a regenerable cell culture from a plant to be transformed, (ii) transforming said culture by bombarding it with DNA-coated microprojectiles,

(iii) identifying or selecting a transformed cell line, and (iv) regenerating a fertile transgenic plant therefrom.

18. The process of claim 17 wherein the cell culture is a friable, embryogenic callus culture.
19. The process of claim 18 wherein the callus culture subjected to bombardment is in clumps of about 30 to 80 mg per clump.
20. The process of claim 19 wherein said callus is initiated on solid media.
21. A plant prepared by the process of claim 17 wherein the plant is regenerated from transformed callus.
22. A transgenic Zea mays plant comprising heterologous DNA which is heritable, wherein said DNA was introduced into said plant, or an ancestor of said plant, at an embryogenic stage.
23. The transgenic Zea mays plant of claim 22 wherein the heterologous DNA was introduced into said plant by microprojectile bombardment.
24. The transgenic Zea mays plant of claim 23 wherein the heterologous DNA was introduced by microprojectile bombardment of a callus culture.

ABSTRACT OF THE DISCLOSURE

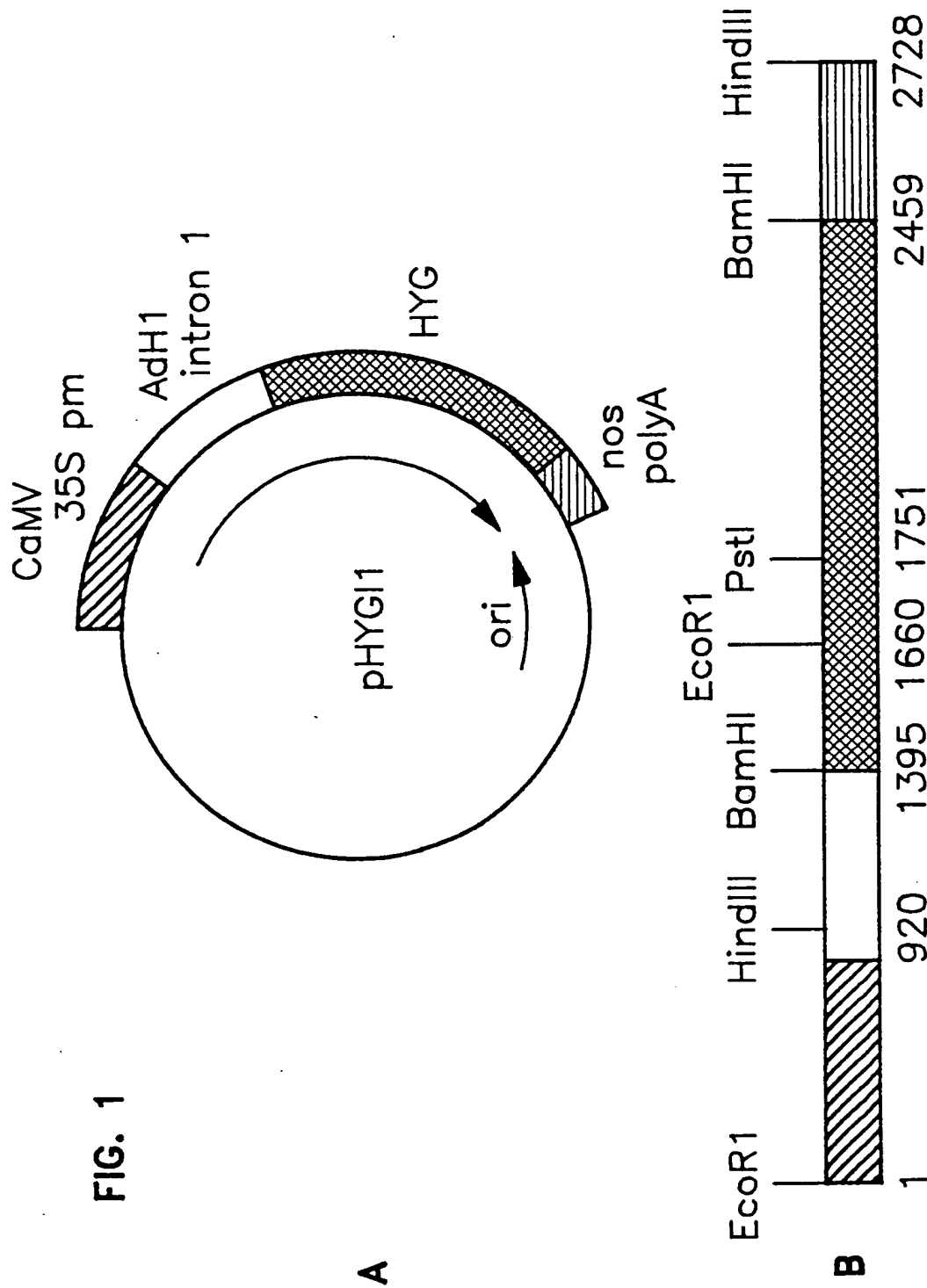
Fertile transgenic Zea mays (corn) plants which stably express heterologous DNA which is heritable are provided along with a process for producing said plants. The preferred process comprises the microprojectile bombardment of friable embryogenic callus from the plant to be transformed. The process may be applicable to other graminaceous cereal plants which have not proven stably transformable by other techniques.

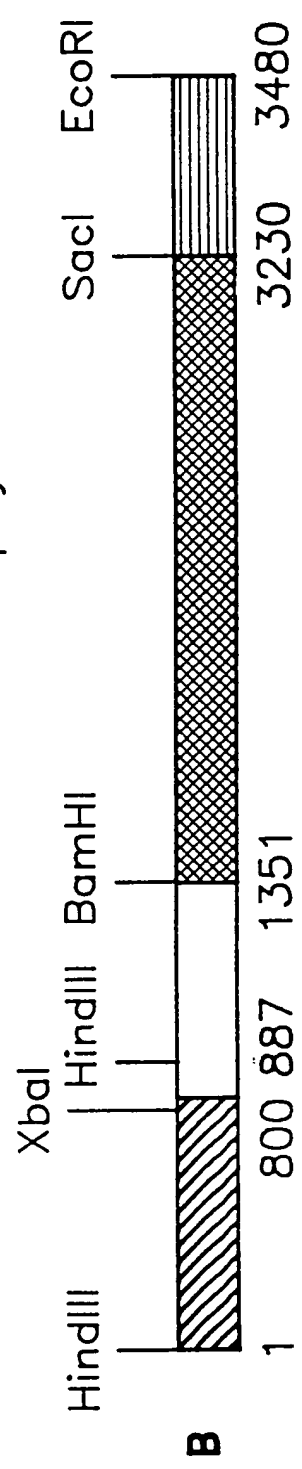
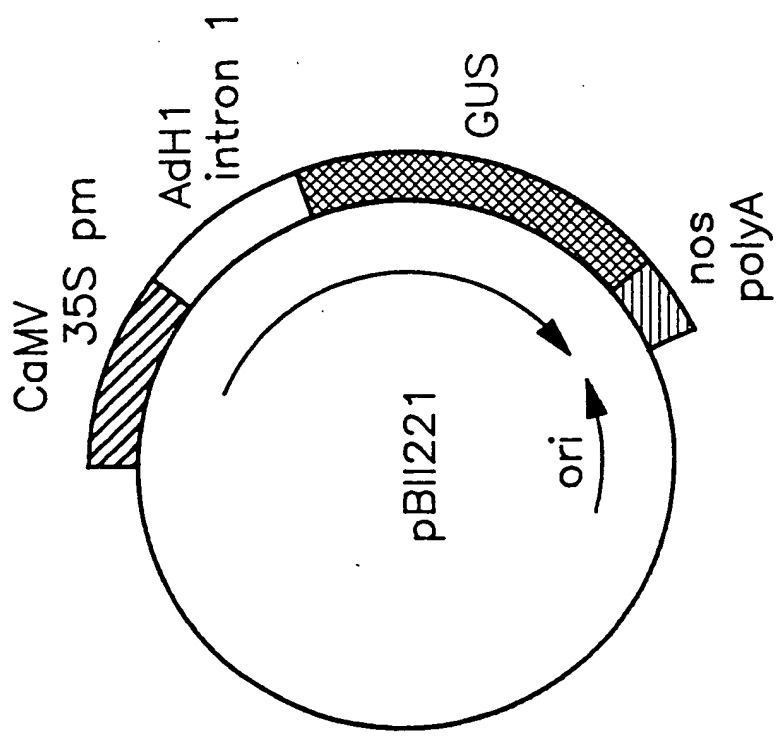
"Express Mail" mailing label number RB 376 455 519  
Date of Deposit April 11, 1990  
I hereby certify that this paper or fee is being deposited  
with the United States Postal Service "Express Mail Post  
Office to Addressee" service under 37 CFR 1.10 on the  
date indicated above and is addressed to the Commis-  
sioner of Patents and Trademarks, Washington, D.C. 20231  
Sheri Ellefson  
printed name  
Sheri Ellefson  
Signature



orig. appln.

FIG. 1





**FIG. 2**

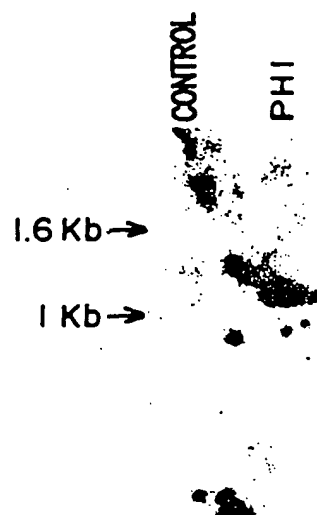
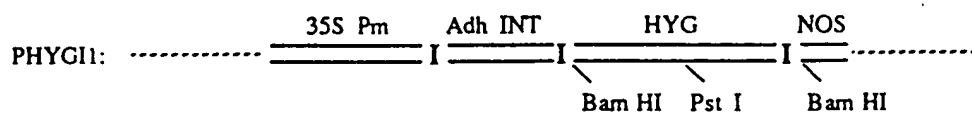
**A**

**B**

FIG. 3

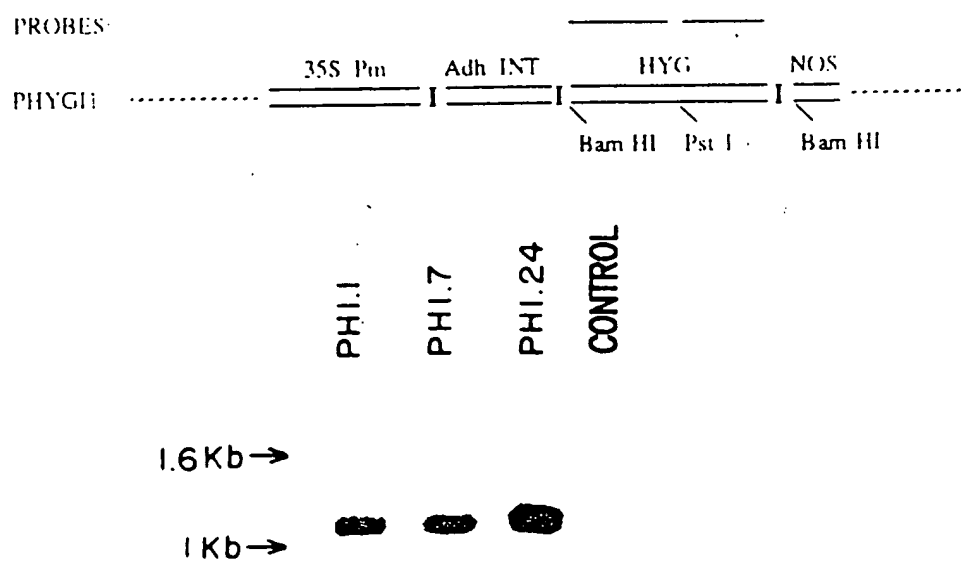
PHI CALLUS

PROBES:



# FIG. 4

## PHI R<sub>0</sub> PLANTS



**FIG. 5**

**PHI R<sub>1</sub> GENERATION**

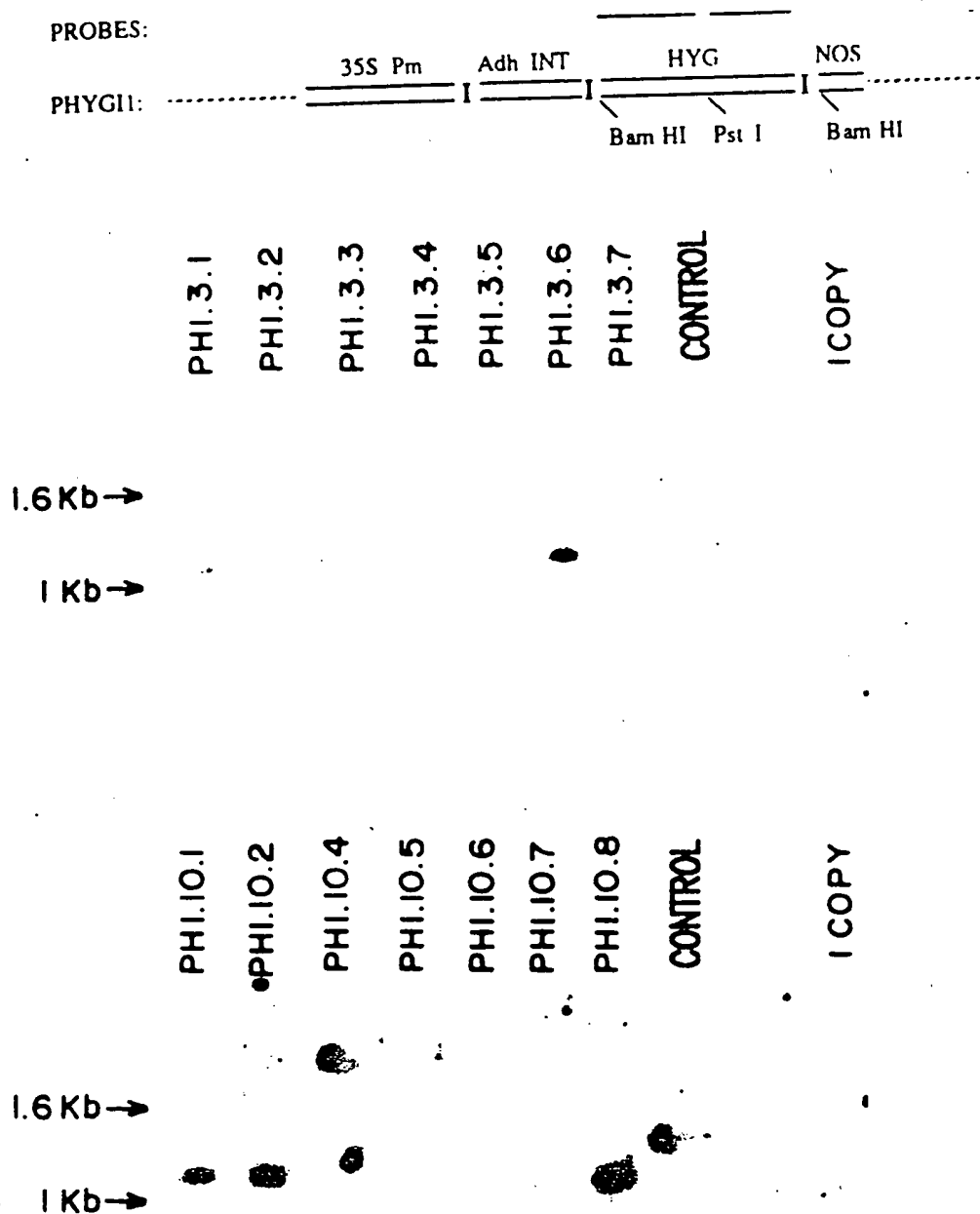
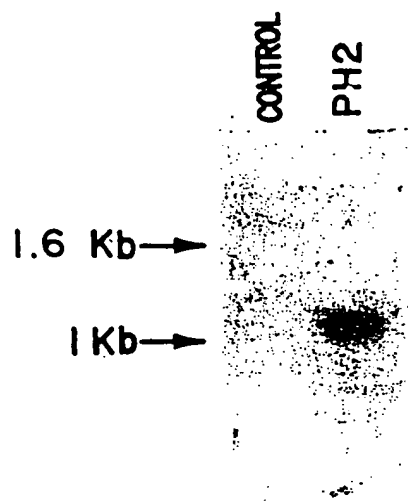
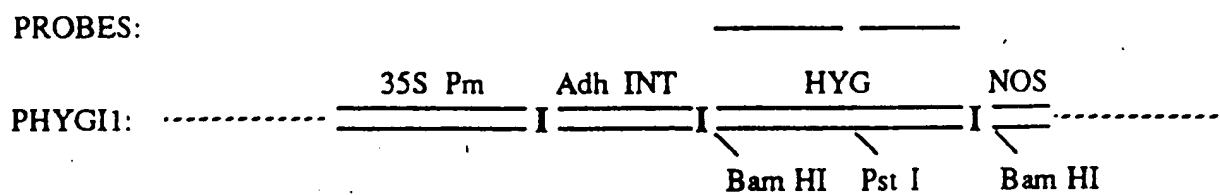


FIG. 6

PH2 CALLUS





US005187073A

**United States Patent** [19]

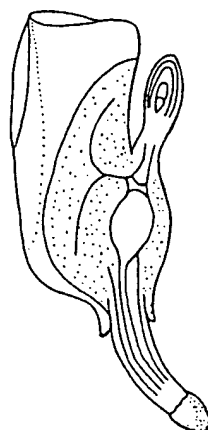
Goldman et al.

[11] **Patent Number:** 5,187,073[45] **Date of Patent:** \* Feb. 16, 1993**[54] PROCESS FOR TRANSFORMING GRAMINEAE AND THE PRODUCTS THEREOF****[75] Inventors:** Stephen L. Goldman, Toledo; Anne C. F. Graves, Bowling Green, both of Ohio**[73] Assignee:** The University of Toledo, Toledo, Ohio**[\*] Notice:** The portion of the term of this patent subsequent to Jan. 5, 2010 has been disclaimed.**[21] Appl. No.:** 436,187**[22] Filed:** Nov. 13, 1989**Related U.S. Application Data****[63]** Continuation of Ser. No. 67,902, Jun. 29, 1987, abandoned, which is a continuation-in-part of Ser. No. 880,271, Jun. 30, 1986, abandoned.**[51] Int. Cl.<sup>5</sup>** ..... C12N 15/00; C12R 1/41**[52] U.S. Cl.** ..... 435/172.3; 435/252.2; 435/320.1; 935/67**[58] Field of Search** ..... 435/172.3, 320.1, 252.2; 935/29, 30, 35, 64, 67**[56] References Cited****PUBLICATIONS**Abbe & Stein, *Amer. J. Bot.*, 41, 285 (1954).Abel, Nelson, De, Hoffman, Rogers, Fraley and Beachy, *Science*, 232, 738 (1986).M. Bagdasarian, Lury, Ruckert, Franklin, M. M. Bagdasarian, Frey, Timmis, *Gene*, 16, 237 (1981).Bevan, Barnes, and Chilton, *Nucleic Acid. Res.*, 11, 369 (1983).Boocock and Coggins, *FEBS Letters*, 154, 127 (1983).Brar, Rambold, Gamborg and Constabel, *Z. Pflanzenphysiol Bd.*, 95, S. 377 (1979).Casse, Boucher, Julliot, Michel and Denaire, *J. Gen. Microbiology*, 113, 229, (1979).Chilton, Saiki, Yadav, Gordon and Quetier, *Proc. Natl. Acad. Sci.*, 77, 4060 (1980).Coe and Sarkar, *Crop Sci.*, 6, 432 (1966).Comai, Facciotti, Giatt, Thompson, Rose and Stalker, *Nature*, 317, 741 (1985).De Cleene and De Ley, *The Botanical Review*, 42, 389 (1976).DeGreve, Dhaese, Seurinck, Lemmers, Montagu and Schell, *J. Mol. Appl. Genet.*, 1, 499 (1983).Depicker, Stachel, Dhaese, Sambryski and Goodman, *J. Mol. Appl. Genet.*, 1, 561 (1983).

(List continued on next page.)

**Primary Examiner**—David T. Fox**Attorney, Agent, or Firm**—Marshall & Melhorn**[57]****ABSTRACT**

A method of producing transformed Gramineae comprising making a wound in a seedling in an area of the seedling containing rapidly dividing cells and inoculating the wound with vir<sup>+</sup> *Agrobacterium tumefaciens*. Also, this same method wherein the vir<sup>+</sup> *A. tumefaciens* contains a vector comprising genetically-engineered T-DNA. There are further provided a transformed pollen grain of a Gramineae, a pollen grain of a Gramineae produced by a plant grown from a seedling infected with vir<sup>+</sup> *A. tumefaciens*, a pollen grain of a Gramineae produced by a plant grown from a seedling infected with vir<sup>+</sup> *A. tumefaciens* containing a vector comprising genetically-engineered T-DNA, a pollen grain of a Gramineae whose cells contain a segment of T-DNA, and Gramineae derived from each of these pollen grains. There are also provided a transformed Gramineae plant, a transformed Gramineae plant derived from a seedling infected with vir<sup>+</sup> *Agrobacterium tumefaciens*, a transformed Gramineae plant derived from a seedling infected with vir<sup>+</sup> *A. tumefaciens* containing a vector comprising genetically-engineered T-DNA and a Gramineae plant whose cells contain a segment of T-DNA. Finally, there are provided transformed Gramineae derived from seedlings infected with vir<sup>+</sup> *Agrobacterium tumefaciens* and transformed Gramineae derived from seedlings infected with vir<sup>+</sup> *A. tumefaciens* containing a vector comprising genetically-engineered T-DNA.

**4 Claims, 7 Drawing Sheets**



## OTHER PUBLICATIONS

- Depicker, De Wilde, G. De Vos, R. De Vos, Montagu and Schell, *Plasmid*, 3, 193 (1980).
- Ditta, Stanfield, Corbin and Helinski, *Proc. Natl. Acad. Sci.*, 77, 7347 (1980).
- Duncan, Williams, Zehr and Widholm, *Planta*, 165, 322 (1985).
- Fromm, Taylor and Walbot, *Nature*, 319, 791 (1986).
- Graves, Ph.D. Dissertation, Bowling Green State University (1982).
- Hain, Stabel, Czernilofsky, Steinbiss, Herrera-Estrella and Schell, *Molec. Gen. Genet.*, 199, 161 (1985).
- Hernalsteens, Thia-Toong, Schell and Montagu, *The EMBO J* 3, 3039 (1984).
- Hooykaas-Van Slogteren, Hooykaas and Schilperoort, *Nature*, 311, 763 (1984).
- Kao, Perry and Kado, *Mol. Gen. Genet.*, 188, 425 (1982).
- Krens, Molendijk, Wullems and Schilperoort, *Nature*, 296, 72 (1982).
- Lin and Kado, *Can. J. Microbiol.*, 23, 1554 (1977).
- Lippincott, Webb and Lippincott, *J. Bact.*, 90, 1955 (1965).
- Lorz, Baker and Schell, *Mol. Gen. Genet.*, 199, 178 (1985).
- Lundquist, Close and Kado, *Mol. Gen. Genet.*, 193, 1 (1984).
- Marton, Wullems, Molendijk and Schilperoort, *Nature*, 277, 129 (1979).
- Matthysse, *J. Bact.*, 154, 906 (1983).
- Murashige and Skoog, *Physiol. Plant.*, 15, 473 (1962).
- Nester, Gordon Amasino and Yanovsky, *Ann. Rev. Plant Physiol.*, 35, 387 (1984).
- Otten, Vreugdenhil and Schilperoort, *Biochem. Biophys. Acta.*, 485, 263 (1977).
- Otten and Schilperoort, *Biochem. Biophys. Acta.*, 527, 497 (1978).
- Otten, DeGreve, Hernalsteens, Montagu, Schieder, Straub and Schell, *Molec. Gen. Genet.*, 183, 209 (1981).
- Paszkowski, Shillito, Saul, Mandak, T. Hohn, B. Hohn and Potrykus, *The EMBO J.*, 3, 2717 (1984).
- Potrykus, Harms and Lorz, *Theor. Appl. Genet.*, 54, 209 (1979).
- Potrykus, Saul, Petruska, Paszkowski and Shillito, *Mol. Gen. Genet.*, 199, 183 (1985).
- Rao, B. Lippincott and J. Lippincott, *Physiol. Plant.*, 56, 374 (1982).
- Sass, in *Corn and Corn Improvement* 63-87 (G. F. Sprague, ed., 1955).
- Sciaky, Montoya and Chilton, *Plasmid*, 1, 238 (1978).
- Schafer, Gory, Kahl, *Nature*, 327, 529 (1987).
- Stalker, Hiatt & Comai, *The J. Biol. Chem.*, 260, 4724 (1985).
- Stonier, *J. Bacteriol.*, 79, 889 (1959).
- Thomashow, Nutter, Postle, Chilton, Blattner, Powell, Gordon and Nester, *Proc. Natl. Acad. Sci.*, 77, 6448 (1980).
- Toole, *Amer. Jour. Bot.*, 11, 325 (1924).
- Turgeon, Wood and Braun, *Proc. Natl. Acad. Sci.*, 73, 3562 (1976).
- Waldron, Murphy, Roberts, Gustafson, Armour and Malcolm, *Plant Molecular Biology*, 5, 103 (1985).
- J. T. Watson, J. Tooze & D. T. Kurtz, *Recombinant DNA* 164-175 (1983).
- Yadav, Postle, Saiki, Thomashow and Chilton, *Nature*, 287, 458 (1980).
- Ohta, *PNAS*, 83, 715-19 (1986).
- Korohoda et al., *Z. Pflanzenphysiol. Bol.*, 94, 95-99 (1979).
- Morgun et al., *Mol. Biol (Kiev)*, 18, 50-53 (1977).
- Abdullah, Cocking and Thompson, *Bio/Technology*, 4, 1087 (1986).
- An, Watson, Stachel, Gordon, and Nester, *The EMBO Journal*, 4: 277-284 (1985).
- Buchanan-Wollaston, Passiatore and Cannon, *Nature*, 328, 172 (1987).
- Byrne, McDonnell, Wright and Carnes, *Plant Cell, Tissue and Organ Culture*, 8, 3-15 (1987).
- Bytebier, Deboeck, DeGreve, Van Montagu, and Hernalsteens, *Proc. Nat'l Acad. Sci USA*, 84, 5345-5349 (1987).
- Chee, Klassy & Slightom, *Gene*, 41, 47-57 (1986).
- Crossway, Oakes, Irvine, Ward, Knauf and Shewmaker, (1986) *Mol. Gen. Genet.*, 202, 179-185.

(List continued on next page.)

## OTHER PUBLICATIONS

- De Cleene, *Phytopath Z.* 113, 81-89 (1985).  
Facciotti, O'Neal, Lee and Shewmaker, *Bio/Technology* 3, 241 (1985).  
Feldmann, and Marks, *Mol. Gen. Genet.* 208, 1-9 (1987).  
Fraley, Rogers, Horsch, Sanders, Flick, Adams, Bittner, Brand, Fink, Fry, Galluppi, Goldberg, Hoffmann and Woo, *Proc. Natl. Acad. Sci., USA*, 80: 4803-7 (1983).  
Fromm, Taylor and Walbot, *Proc. Natl. Acad. Sci. U.S.A.*, 82, 5824-5828 (1985).  
Graves and Goldman, Abstract, *Proceedings of the Crown Gall Conference*, p. 20 Sep. 1986.  
Graves and Goldman, *Plant. Mol. Biol.*, 7, 43 (1986).  
Grimsley, T. Hohn, Davies, and B. Hohn, *Nature*, 325, 177 (1987).  
Grimsley, Ramos, Hein, and B. Hohn, *Bio/Technology*, 6, 185 (1988).  
Kado and Liu, *Journal of Bacteriology*, 145, 1365 (1981).  
Klein, Wolf, Wu and Sanford, *Nature*, 327, 70, (1987).  
Kyoizuka, Hayashi and Shimamoto, *Mol. Gen. Genet.*, 206, 408-413 (1987).  
J. Lippincott, Margot and B. Lippincott, *J. Bact.*, 132, 824 (1977).  
Murai et al., *Science*, 222, 476 (1983).  
Owens and Cress, *Plant Physiol.*, 77, 87-94 (1985).  
Slightom, Sun and Hall, *Proc. Nat'l Acad. Sci. USA*, 80, 1987 (1983).  
Yamada, Zhi-Qi and Ding-Tai, *Plant Cell Reports*, 5, 85 (1986).  
Graves and Goldman, *Journal of Bacteriology*, 169, 1745 (1987).  
Korohoda and Strzalka, *Z. Pflanzenphysiol. Bd.*, 94, 95-99 (1979).  
Morgun, Chernykh, Kordyum, *Chem. Abstracts*, 88(11), 71618k (1977) (*Mol. Biol. (Kiev)*, 18, 50 (1977)).  
Ohta, *Proc. Natl. Acad. Sci. USA*, 83, 715-19 (1986).  
de la Pena, Lorz and Schell, *Nature*, 325, 274 (1987).  
*Chemistry and Engineering News*, p. 21, Jul. 13, 1987.  
*The Economist*, p. 92, Mar. 7, 1987.

FIG. 1A

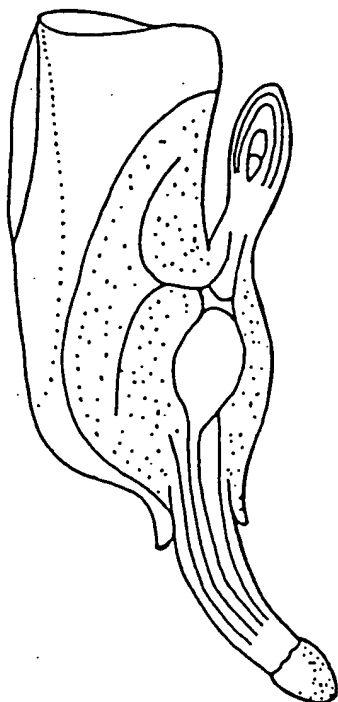


FIG. 1B

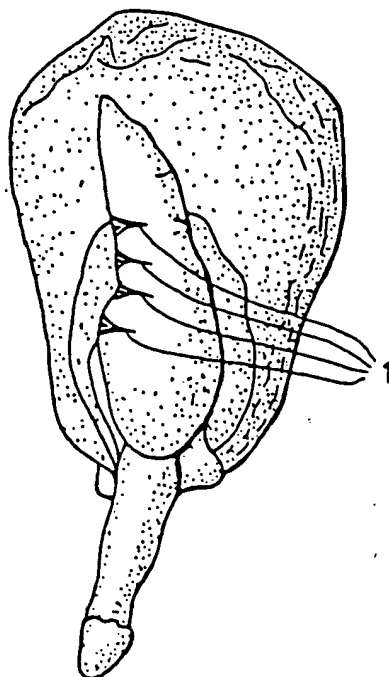


FIG. 2A

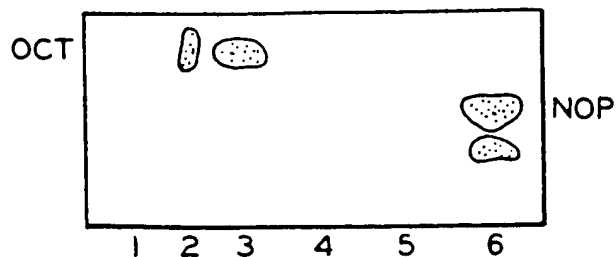


FIG. 2B

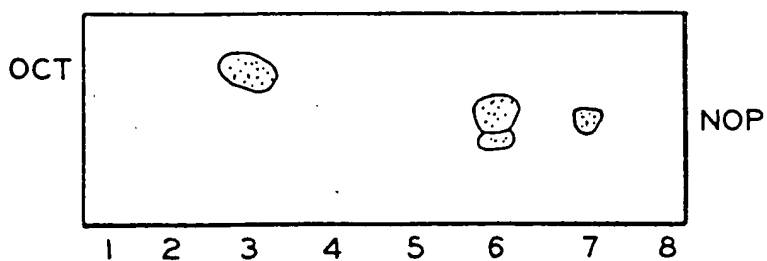


FIG. 2C

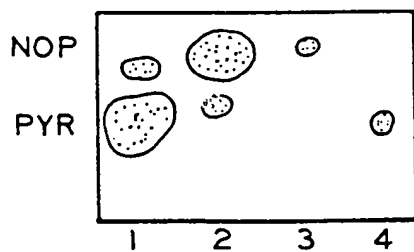


FIG. 2D

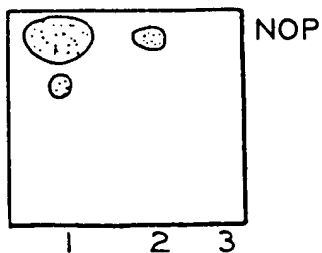


FIG. 3A

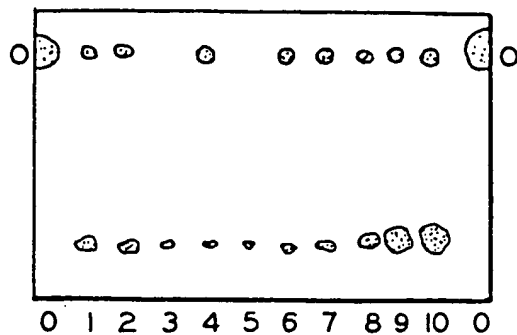


FIG. 3B

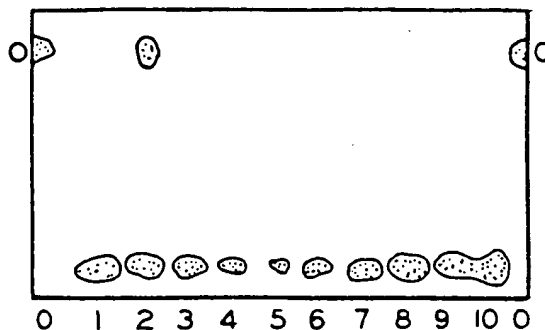


FIG. 3C

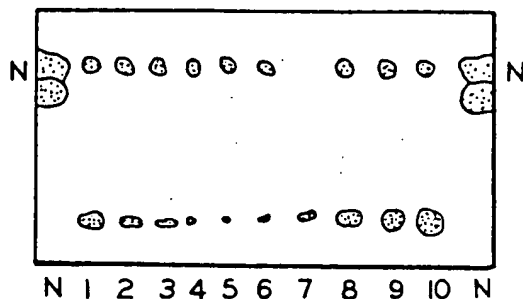


FIG. 3D

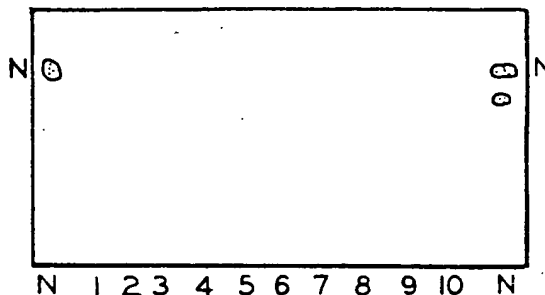


FIG. 4A

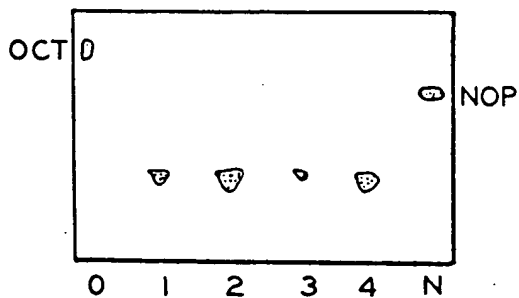


FIG. 4B

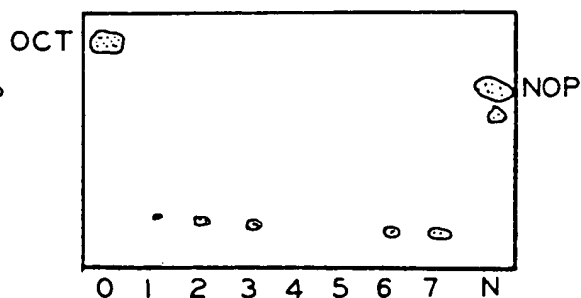


FIG. 4C

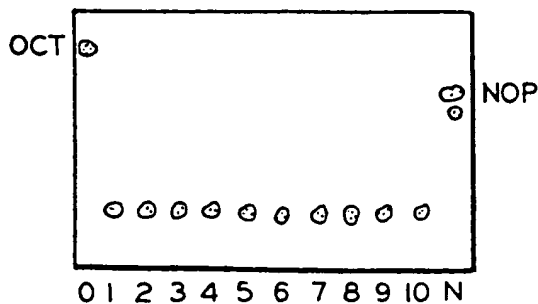


FIG. 5A

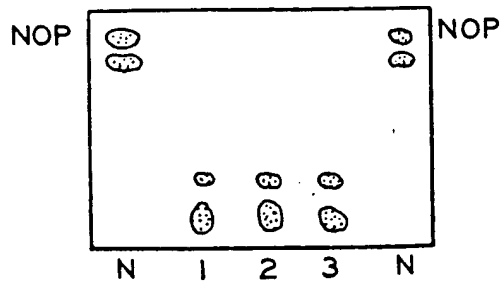


FIG. 5C

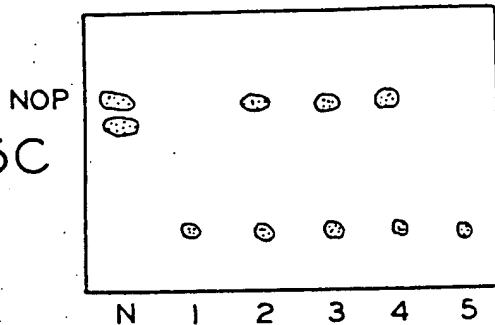


FIG. 5D

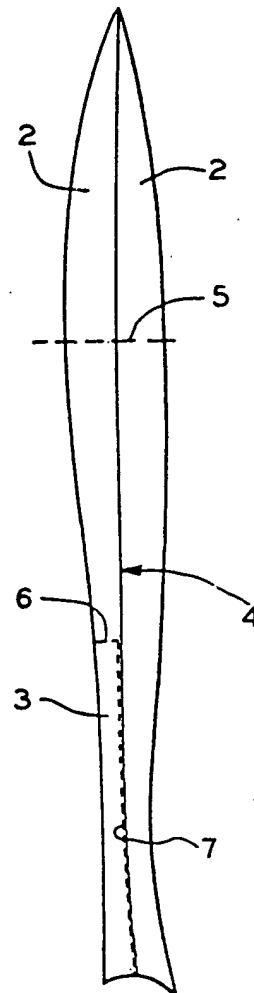
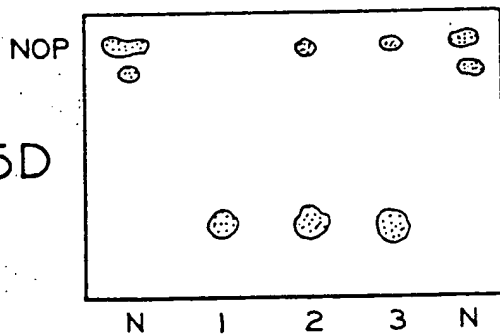


FIG. 5B

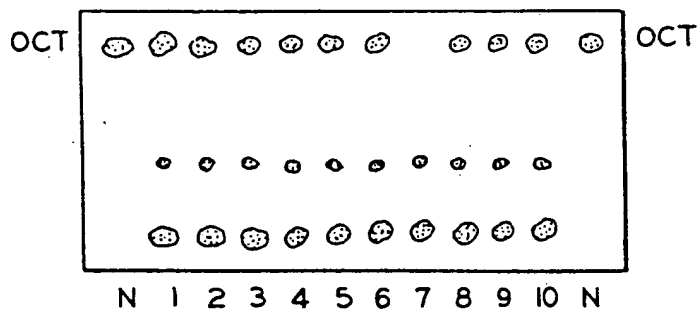


FIG. 6

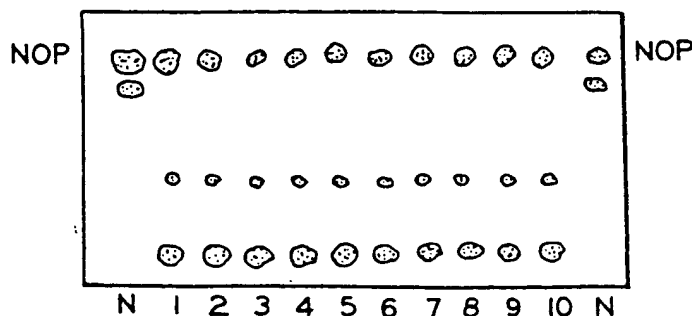


FIG. 7

FIG. 8

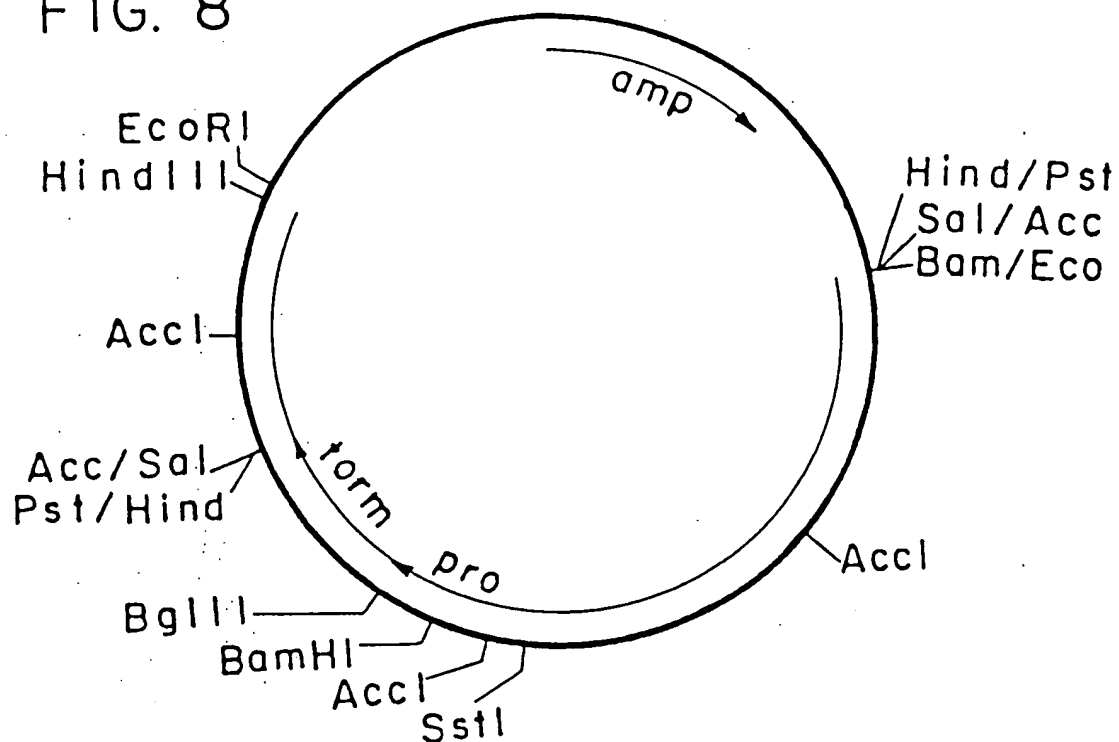


FIG. 9

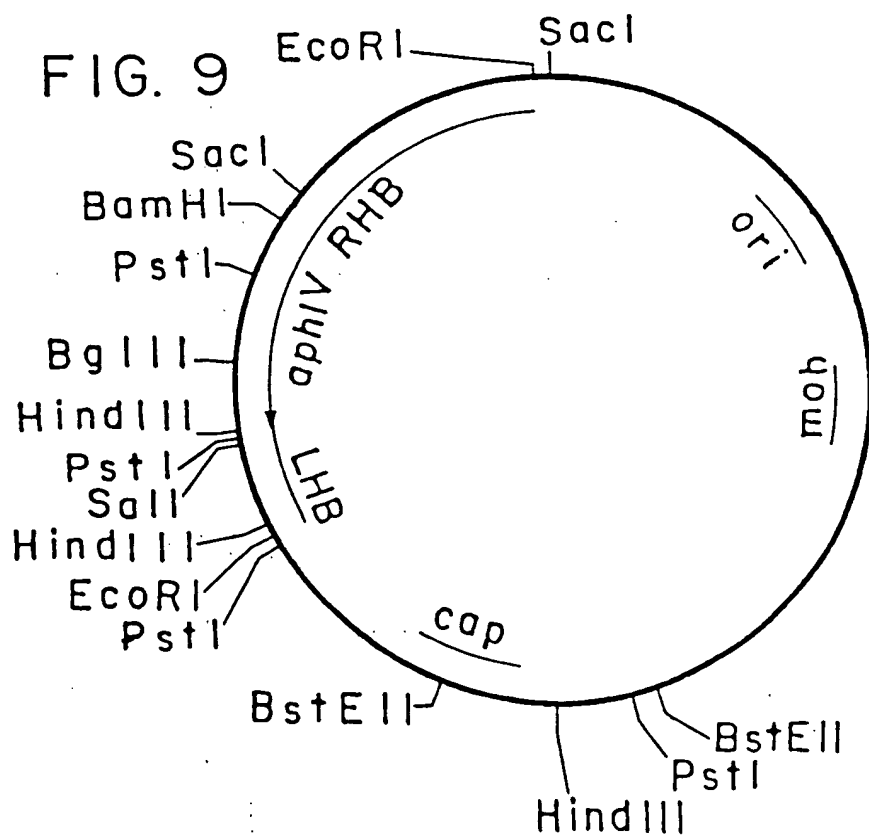


FIG. 10

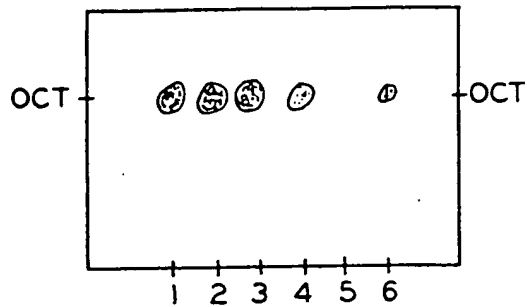


FIG. 11

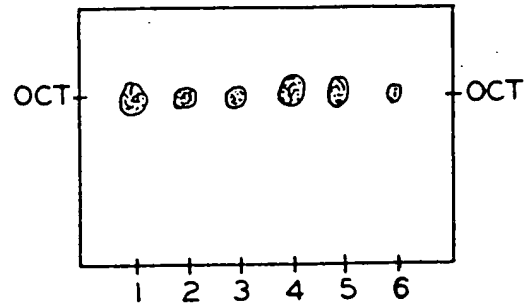


FIG. 12

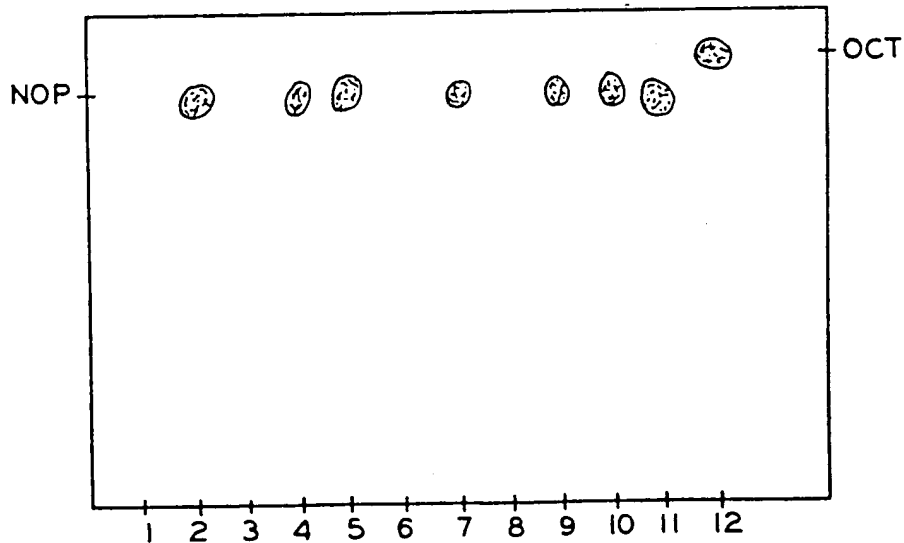




FIG. 13

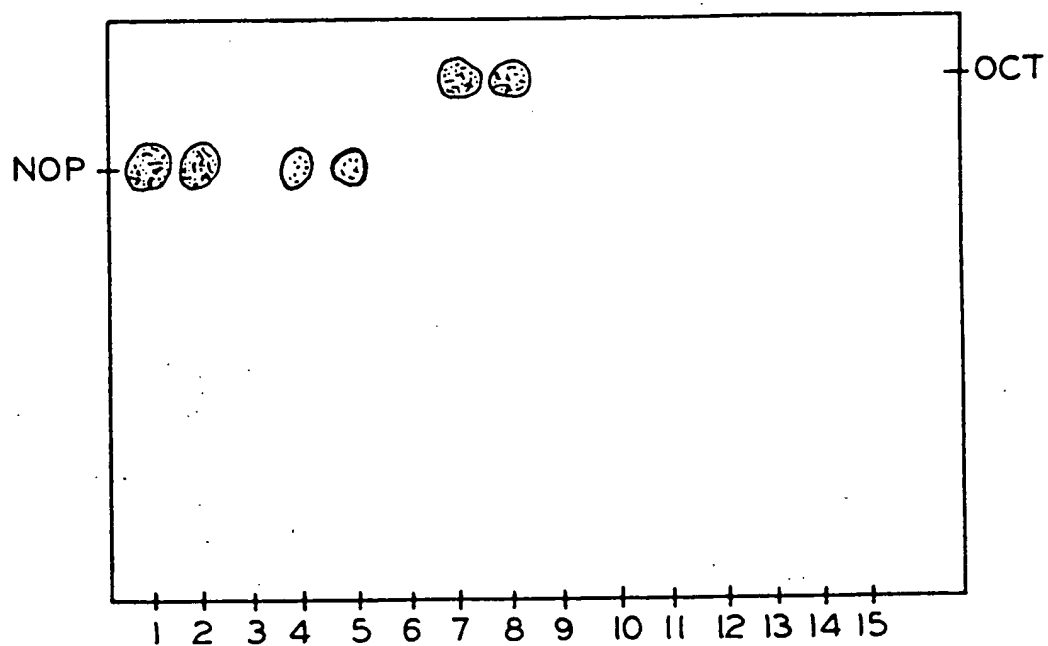
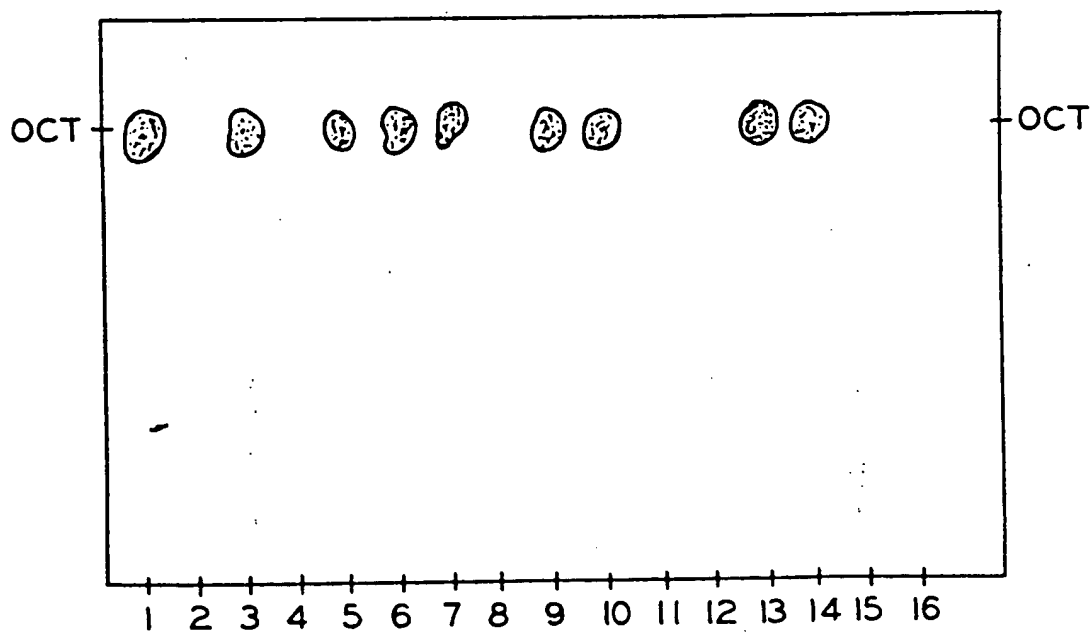


FIG. 14



## PROCESS FOR TRANSFORMING GRAMINEAE AND THE PRODUCTS THEREOF

This is a continuation of application Ser. No. 07/067,902, filed Jun. 29, 1987, now abandoned, which was a continuation-in-part of application Ser. No. 06/880,271, filed Jun. 30, 1986, now abandoned.

### BACKGROUND OF THE INVENTION

Virulent strains of the soil bacterium *Agrobacterium tumefaciens* are known to infect dicotyledonous plants and to elicit a neoplastic response in these plants. The tumor-inducing agent in the bacterium is a plasmid that functions by transferring some of its DNA into its host plant's cells where it is integrated into the chromosomes of the host plant's cells. This plasmid is called the Ti plasmid, and the virulence of the various strains of *A. tumefaciens* is determined in part by the vir region of the Ti plasmid which is responsible for mobilization and transfer of the T-DNA. The T-DNA section is delimited by two 23-base-pair repeats designated right border and left border, respectively. Any genetic information placed between these two border sequences may be mobilized and delivered to a susceptible host. Once incorporated into a chromosome, the T-DNA genes behave like normal dominant plant genes. They are stably maintained, expressed and sexually transmitted by transformed plants, and they are inherited in normal Mendelian fashion.

The lump of plant tumor tissue that grows in an undifferentiated way at the site of the *A. tumefaciens* infection is called a crown gall. Cells of crown gall tumors induced by *A. tumefaciens* synthesize unusual amino acids called opines. Different strains of *A. tumefaciens* direct the synthesis of different opines by the crown gall cells, and the particular opine induced is a characteristic of the strain which infected the plant. Further, the ability to catabolize the particular opine induced by a given strain is also characteristic of that strain.

Opines are not normally synthesized by *A. tumefaciens* or by the uninfected host plants. Although it is the T-DNA which codes for the enzymes involved in the synthesis of the opines, the opine synthases, these genes are expressed only in infected plant tissue. This type of expression is consistent with the observation that these genes are under the control of eukaryotic regulatory sequences on the T-DNA.

The most common opines are octopine and nopaline. The opine synthase that catalyzes the synthesis of octopine is lysopine dehydrogenase, and the opine synthase that catalyzes the synthesis of nopaline is nopaline dehydrogenase.

When crown gall cells are put into culture, they grow to form a callus culture even in media devoid of the plant hormones that must be added to induce normal plant cells to grow in culture. A callus culture is a disorganized mass of relatively undifferentiated plant cells. This ability of crown gall cells to grow in hormone-free media is also attributable to the presence of the T-DNA in the transformed host plant cells since genes which direct the synthesis of phyto-hormone are also associated with the T-DNA.

A DNA segment foreign to the *A. tumefaciens* and to the host plant which is inserted into the T-DNA by genetic manipulation will also be transferred to host plant's cells by *A. tumefaciens*. Thus, the Ti plasmid can be used as a vector for the genetic engineering of host

plants. Although, in wild type *A. tumefaciens* there is only one Ti plasmid per bacterium, in genetically-engineered *A. tumefaciens*, the vir region and the T-DNA do not have to be carried on the same Ti plasmid for transfer of the T-DNA to occur. The vir region and the T-DNA can be carried on separate plasmids contained within the same *Agrobacterium*.

It has generally been assumed that the host range of *A. tumefaciens* was limited to the dicotyledons, and that transformation of monocotyledons could not be effected with this bacterium. Indeed, no one has reported the transformation of any member of the monocotyledonous Gramineae family by infection with *A. tumefaciens*.

However, recently, Hooykaas-Van Slogteren et al., in *Nature*, 311, 763 (1984), reported the production of small swellings at wound sites infected with *A. tumefaciens* on monocotyledons of the Liliaceae and Amaryllidaceae families. Opines were detected in plant cells taken from the wound sites of the infected plants.

Also, Hernalsteens et al. reported in *The EMBO Journal*, 3, 3039 (1984) that cultured stem fragments of the monocotyledon *Asparagus officinalis*, a member of the family Liliaceae, infected with *A. tumefaciens* strain C58 developed tumorous proliferations. One of these tumorous proliferations could be propagated on hormone-free medium, and opines were detected in the established callus culture derived from this tumorous proliferation.

In 1982, Anne C. F. Graves reported in her Ph.D. Dissertation entitled "Some Tumorigenic Activities of *Agrobacterium Tumefaciens* (Smith and Town) Conn." (Bowling Green State University) that irregular masses of tissue developed on gladiolus discs in response to inoculations with *A. tumefaciens* C58N and B6. These masses of tissue appeared to be the same as, and have cellular morphology similar to, those that develop on potato tuber discs. Gladiolus is a member of the monocotyledonous Iridaceae family. A compound that comigrated with the octopine standard during electrophoresis was found in the proliferations on the gladiolus discs that were induced by strain B6, and one that migrated just behind the octopine standard occurred in those induced by C58N. Also, octopine dehydrogenase was found in extracts of the cellular proliferations induced by *A. tumefaciens* B6 but not in those induced by *A. tumefaciens* C58N.

Dr. Graves also described the response of certain other monocots to inoculation with *A. tumefaciens*. No cellular proliferation was observed on ginger root rhizome discs, and the results with tulip bulb discs were inconclusive. Cellular proliferations on discs of the rhizomes of cattail and skunk cabbage were limited to several layers of clear cells at the ends of vascular bundles in the early spring.

DeCleene and DeLey in *The Botanical Review*, 42, 389 (1976) reported the results of an extensive study of the plant host range of *A. tumefaciens*. Their article teaches that monocots of the orders Liliales and Arales are susceptible to infection with *A. tumefaciens* but that monocotyledons in general are unsusceptible to *A. tumefaciens* infection. In particular, their article reports that the Gramineae tested were not susceptible to infection with *A. tumefaciens*. Susceptibility to *A. tumefaciens* infection was determined by whether a swelling or tumor developed at the wound site.

Lorz et al. in *Mol. Gen. Genet.*, 199, 178 (1985), Fromm et al. in *Nature*, 319, 791 (1986) and Portrykus et al. in *Mol. Gen. Genet.*, 199, 183 (1985) have reported the

transformation of Gramineae by direct gene transfer to protoplasts. Protoplasts are plant cells from which the cell wall has been removed by digestion with enzymes. Lorz et al. transformed protoplasts of *Triticum monococcum* using a DNA construct containing the nopaline synthase promoter and the polyadenylation regulatory signal of the octopine synthase gene. Fromm et al. discloses that the electroporation-mediated transfer of plasmid pCaMVNEO (comprising the cauliflower mosaic virus 35 S promoter, the neomycin phosphotransferase II gene from the transposon Tn5 and the nopaline synthase 3' region) into maize protoplasts results in stably-transformed maize cells that are resistant to kanamycin.

To obtain transformed plants from the transformed cells generated using either the infection techniques of Hooykaas-Van Slogteren et al., Hernalsteens et al., Graves and DeCleene and DeLey or the direct gene transfer techniques of Lorz et al., Fromm et al. and Portrykus et al., the plants would have to be regenerated from protoplasts or single cell cultures. However, no one has yet been able to regenerate plants from protoplasts or single cell cultures of the Gramineae. Indeed, no means of producing transformed plants or other transformed differentiated organs or tissues of the Gramineae is currently known, and no method yet exists for transforming the Gramineae in a manner which allows for the expression of exogenous DNA in agriculturally important forms or parts of the Gramineae such as seeds, pollen, ears or plants.

Finally, PCT International Publication No. WO 86/00931 (Simpson et al.) published Feb. 13, 1986, teaches in vivo methods for transforming and regenerating intact plants. This patent application discloses that the methods of the invention can be used for the transformation of any plant that forms a shooty tumor following infection with an *A. tumefaciens* shooty mutant strain. However, as noted above, the Gramineae are not known to produce tumors or even swellings in response to inoculation with *A. tumefaciens*. In the practice of the present invention, no tumors, swellings or cellular proliferations of any kind have been observed on the Gramineae in response to inoculation with *A. tumefaciens*.

### SUMMARY OF THE INVENTION

According to the present invention, there is now provided a method of producing transformed Gramineae comprising making a wound in a seedling in an area of the seedling containing rapidly dividing cells and inoculating the wound with vir<sup>+</sup> *A. tumefaciens*. In the preferred practice of the invention, the wound is made in an area of the seedling which gives rise to the germ line cells. Also, preferred is the use of vir<sup>+</sup> *A. tumefaciens* which contains a vector comprising genetically-engineered T-DNA.

According to another aspect of the invention, there are provided: (1) transformed pollen grains; (2) a transformed pollen grain produced by a plant grown from a seedling infected with vir<sup>+</sup> *A. tumefaciens*; (3) a transformed pollen grain produced by a plant grown from a seedling infected with vir<sup>+</sup> *A. tumefaciens* which contains a vector comprising genetically-engineered T-DNA; (4) a pollen grain whose cells contain at least a segment of T-DNA; and (5) Gramineae derived from each of these four pollen grains. In addition, there are provided: (1) a transformed Gramineae plant; (2) a transformed Gramineae plant derived from a seedling

infected with vir<sup>+</sup> *A. tumefaciens*; (3) a transformed Gramineae plant derived from a seedling infected with vir<sup>+</sup> *A. tumefaciens* which contains a vector comprising genetically-engineered T-DNA; and (4) a Gramineae plant whose cells contain a segment of T-DNA. Finally, there are provided transformed Gramineae derived from seedlings infected with vir<sup>+</sup> *A. tumefaciens* and transformed Gramineae derived from seedlings infected with vir<sup>+</sup> *A. tumefaciens* which contains a vector comprising genetically-engineered T-DNA.

The invention is clearly useful since it provides, for the first time, a method for transforming Gramineae which results in the production of transformed differentiated organs and tissue such as leaves, plants and pollen. Thus, the invention provides, for the first time, a method of transforming Gramineae which allows for the expression of exogenous DNA in agriculturally important forms or parts of the Gramineae. Many of the Gramineae (such as corn, oats, rye, barley, sorghum, rice and wheat) are, of course, commercially important food sources for humans and other animals, and the invention allows for the development of strains of Gramineae having altered or superior traits, such as higher yielding strains and strains having resistance to herbicides or better nutritional value, by providing a means whereby exogenous DNA coding for such traits can be incorporated into the Gramineae.

### BRIEF DESCRIPTION OF THE DRAWINGS

FIGS. 1-18 show a corn seedling which is ninety-six hours old. FIG. 1A is the side view of the seedling, and FIG. 1B is the front view of the same seedling.

FIG. 2A is a drawing of developed paper electrophoretogram which shows the results of the electrophoresis of the products produced by incubating cell-free extracts of yellow Iochief corn seedlings inoculated with *A. tumefaciens* strain B6 with a reaction medium containing reagents which allow for the detection of lysopine dehydrogenase enzyme activity (lysopine dehydrogenase reaction medium). Certain controls were also electrophoresed.

FIG. 2B is a drawing of a developed paper electrophoretogram which shows the results of the electrophoresis of the products produced by incubating cell-free extracts of yellow Iochief corn seedlings with a reaction medium containing reagents which allow for the detection of nopaline dehydrogenase enzyme activity (nopaline dehydrogenase reaction medium). Certain controls were also electrophoresed.

FIG. 2C is a drawing of a developed paper electrophoretogram which shows the presence or absence of pyronopaline in various materials.

FIG. 2D is a drawing of a developed paper electrophoretogram. The materials electrophoresed were produced by the catabolism by certain strains of *A. tumefaciens* of the products produced by incubating cell-free extracts of yellow Iochief corn seedlings inoculated with *A. tumefaciens* strain C58 with nopaline dehydrogenase reaction medium. Certain controls were also electrophoresed.

FIG. 3A is a drawing of a developed paper electrophoretogram which shows the results of the electrophoresis of the products produced by incubating cell-free extracts of single B6-inoculated yellow Iochief corn seedlings with lysopine dehydrogenase reaction medium.

FIG. 3B is a drawing of a developed paper electrophoretogram which shows the results of the electrophoresis of the products produced by incubating cell-free extracts of single B6-inoculated yellow Iochief corn seedlings with lysopine dehydrogenase reaction medium.

resis of the products produced by incubating cell-free extracts of single yellow lochief corn seedlings inoculated with either *A. tumefaciens* strain A348 or strain 238MX with lysopine dehydrogenase reaction medium.

FIG. 3C is a drawing of a developed paper electrophoretogram showing the results of the electrophoresis of the products produced by incubating cell-free extracts of single C58-inoculated yellow lochief corn seedlings with nopaline dehydrogenase reaction medium.

FIG. 3D is a drawing of a developed paper electrophoretogram showing the results of the electrophoresis of the products produced by incubating cell-free extracts of single yellow lochief corn seedlings inoculated with *A. tumefaciens* strain JK 195 with nopaline dehydrogenase reaction medium.

FIG. 4A is a drawing of a developed paper electrophoretogram which is the result of the electrophoresis of the cell-free sonicates of *A. tumefaciens* strains B6 and C58.

FIG. 4B is a drawing of a developed paper electrophoretogram showing the results of the electrophoresis of the products produced by incubating the cell-free sonicates of *A. tumefaciens* strains B6 and C58 for four hours with an appropriate reaction medium. Also electrophoresed were the reaction media alone.

FIG. 4C is a drawing of a developed paper electrophoretogram which shows the results of the electrophoresis of the product produced by incubating the cell-free extracts of single uninfected seedlings of the yellow lochief strain of corn for four hours with lysopine dehydrogenase reaction medium or with nopaline dehydrogenase reaction medium.

FIG. 5A is a drawing of a developed electrophoretogram showing the results of the electrophoresis of the products produced by incubating the cell-free extracts of single embryonic leaves from plants grown from C58-inoculated yellow lochief corn seedlings with nopaline dehydrogenase reaction medium.

FIG. 5B is a drawing of a corn leaf of meristematic origin showing the areas dissected for assay for nopaline dehydrogenase activity.

FIG. 5C is a drawing of a developed electrophoretogram showing the results of the electrophoresis of the products produced by incubating the cell-free extracts of the dissected sections of four meristematic leaves from four separate plants grown from C58-inoculated yellow lochief corn seedlings with nopaline dehydrogenase reaction medium.

FIG. 5D is a drawing of a developed electrophoretogram showing the results of the electrophoresis of the products produced by incubating cell-free extracts of pollen from individual plants grown from C58-inoculated yellow lochief corn seedlings with nopaline dehydrogenase reaction medium.

FIG. 6 is a drawing of a developed electrophoretogram which is the result of the electrophoresis of the products produced by incubating the cell-free extracts of single seedlings of yellow lochief corn inoculated with *A. tumefaciens* strain CA19 with lysopine dehydrogenase reaction medium.

FIG. 7 is a drawing of a developed electrophoretogram which shows the results of the electrophoresis of the products produced by incubating the cell-free extracts of single C58-inoculated seedlings of strain PA91 corn with nopaline dehydrogenase reaction medium.

FIG. 8 is a restriction site and function map of plasmid pCEL30.

FIG. 9 is a restriction site and function map of plasmid pCEL44.

FIG. 10 is a drawing of developed paper electrophoretogram which shows the results of the electrophoresis of the products produced by incubating cell-free extracts of single B6-inoculated rye seedlings with lysopine dehydrogenase reaction medium.

FIG. 11 is a drawing of a developed paper electrophoretogram which shows the results of the electrophoresis of the products produced by incubating cell-free extracts of single B6-inoculated barley seedlings with lysopine dehydrogenase reaction medium.

FIG. 12 is a drawing of a developed paper electrophoretogram showing the results of the electrophoresis of the products produced by incubating cell-free extracts of single C58-inoculated oat seedlings with nopaline dehydrogenase reaction medium.

FIG. 13 is a drawing of a developed electrophoretogram which is the result of the electrophoresis of the products produced by incubating the cell-free extracts of single C58-inoculated wheat seedlings with nopaline dehydrogenase reaction medium and with lysopine dehydrogenase reaction medium.

FIG. 14 is a drawing of a developed electrophoretogram showing assays for enzymatic activity in the upper leaves of transformed plants.

In all of the drawings where they are used, the designations "O" and "OCT" mean octopine, and the designations "N" or "NOP" mean nopaline. These designations are used on the drawings to refer to the lane of the electrophoretogram containing the synthetic octopine or nopaline standard or to show the location of spot formed by the synthetic octopine or nopaline standard after electrophoresis.

In some of the drawings there are spots which do not co-migrate with either the synthetic octopine or nopaline standards. These spots are formed by unreacted reactants in reaction media or are formed by naturally-occurring materials found in the cell-free extracts of the corn seedlings.

#### DETAILED DESCRIPTION OF THE PRESENTLY PREFERRED EMBODIMENTS

According to the invention, there is provided a method of producing transformed Gramineae comprising making a wound in a seedling in an area of the seedling containing rapidly dividing cells and inoculating the wound with vir+ *A. tumefaciens*. The term Gramineae, as used herein, is meant to include all of the forms and parts of the Gramineae, including plants, seeds, seedlings, pollen, kernels, ears, leaves, stalks and embryos. Similarly, the terms "corn", "oats", "wheat", "rye" and "barley" are meant to include all forms and parts of the corn, oats, wheat, rye or barley. "Transformed" is used herein to mean genetically modified by the incorporation of exogenous DNA into cells. "Exogenous DNA" is DNA not normally found in the strain of Gramineae which is to be transformed. Exogenous DNA may be obtained from prokaryotic or eukaryotic sources, including strains of Gramineae other than the one to be transformed.

In practicing the method of the invention, the strain of Gramineae to be transformed is sterilized and is then germinated until the radicle (primary root) and stem emerge from the seed. This occurs after about four days of germination, but this time may be shortened by first soaking the seeds.

The wound is preferably made in an area of rapidly dividing cells which gives rise to germ line cells. After making the wound, the seedling is inoculated by dripping a solution of vir<sup>+</sup> *A. tumefaciens* into the wound. Vir<sup>+</sup> *A. tumefaciens* are bacteria which are capable of mobilizing and transferring T-DNA into host plant cells, and an *A. tumefaciens* carrying a plasmid, whether natural or engineered, coding for these functions is vir<sup>+</sup>. Thus, a strain of *A. tumefaciens* that carries a wild-type Ti plasmid is vir<sup>+</sup> and can be used in the present invention. Many such strains are known and are publicly available. See e.g., American Type Culture Collection Catalogue (ATCC) of Strain I, p. 66 (15th edition, 1982). The vir<sup>+</sup> region of a wild type Ti plasmid can be used to mobilize and transfer T-DNA on the same Ti plasmid or to deliver T-DNA on another plasmid contained in the same bacterium. In addition, the mobilization and transfer functions can be supplied by helper plasmids. Such helper plasmids have been described by Ditta et al. in PNAS, 77, 7347 (1980) and by Bagdasarian et al. in *Gene*, 16, 237 (1981). Thus, a strain of *A. tumefaciens* that carries a helper plasmid is also vir<sup>+</sup>. Finally, the mobilization and transfer functions may be coded on the same engineered plasmid which contains the T-DNA, and bacteria containing such a plasmid are also vir<sup>+</sup>.

The T-DNA transferred by the vir<sup>+</sup> *A. tumefaciens* may be native T-DNA or may preferably be genetically-engineered T-DNA. Genetically-engineered T-DNA is a DNA construct comprising T-DNA border sequences, a heterologous gene and a transcription unit connected in operable order. Methods of preparing such constructs are known in the art.

A heterologous gene is a gene which is not normally found in the T-DNA and which is also not normally found in the DNA of the strain of Gramineae which is to be transformed. Heterologous genes may be isolated from prokaryotic and eukaryotic sources, including strains of Gramineae other than the one to be transformed. Of particular interest are those heterologous genes which confer agronomically significant traits on plants containing them.

The heterologous gene is flanked by a transcription unit containing, e.g., promoters and terminators, which allow for expression of the heterologous gene in the strain of Gramineae to be transformed. The heterologous gene-transcription-unit construct is flanked by the border sequences. Any T-DNA border sequence, native or synthesized, can be used to flank the heterologous-gene-transcription-unit construct as long as the border sequence functions to integrate the heterologous gene into the cell genome of the strain of Gramineae to be transformed. The genetically-engineered T-DNA is linked to a DNA fragment containing a replicon that is functional in *Agrobacterium* to form a vector.

After the seedlings are inoculated with the vir<sup>+</sup> *A. tumefaciens*, they are incubated until transformation has taken place at which time the seedlings are planted and allowed to grow at least until such time as they have produced pollen. It is interesting to note that using the method of the invention, no tumorous growth of any kind, including crown galls, calli or tumorous overgrowths, has been observed, even on the original inoculated seedling.

By inoculating the seedlings in the preferred area, transformation of pollen has been achieved. The resulting transformed pollen can be used to fertilize transformed and untransformed plants. The embryos can be

excised from the resultant progeny ears and grown to produce transformed plants. Alternatively, of course, the plants resulting from this mating can be allowed to produce seeds which are, in turn, used to grow transformed whole plants which produce another crop of seeds. Thus, future generations can be derived from the original transformed seedling by sexual reproduction. Those skilled in the art will recognize that various and numerous progeny carrying the trait coded for by the exogenous DNA can be produced using known breeding techniques.

## TRANSFORMATION OF CORN

### EXAMPLE I

#### A. Preparation of Bacteria

A single colony of the *A. tumefaciens* strain B6 was inoculated into a yeast extract broth (YEB) containing 0.1% yeast extract, 0.8% nutrient broth and 0.5% sucrose dissolved in water. The yeast extract and nutrient broth were purchased from Difco Laboratories, Detroit, Mich. The sucrose was purchased from either Fisher Scientific, Detroit, Mich., or Sigma, St. Louis, Mo. The bacteria were incubated in the YEB for 48 hours at 27° C. in a shaking water bath or until such time as they had reached a final concentration of  $3.8 \times 10^9$  cells per milliliter.

The B6 strain is a standard wild type strain of *A. tumefaciens*. It is virulent (vir<sup>+</sup>), and it codes for the production of lysopine dehydrogenase in suitable plant hosts. It was obtained from James and Barbara Lippincott, Northwestern University, Evanston, Ill. Some of its properties have been described in Stonier, *J. Bact.*, 79, 889 (1960). The B6 strain is also on deposit at the American Type Culture Collection (ATCC), Rockville, Md., and it has been given accession number 23308.

#### B. Preparation of Corn

Seeds of the inbred yellow Iochief strain of corn were obtained from The Andersons, Maumee, Ohio, or from Botzum, 43 East Market, Akron, Ohio. This strain is a standard strain which is available commercially.

The yellow Iochief seeds were sterilized using the following procedure. The seeds were first immersed for two minutes in a solution containing 7 parts 95% ethanol and 2.5 parts distilled water. Next, the seeds were incubated for five minutes in a solution of 0.5% (weight/volume) HgCl<sub>2</sub> in distilled water. Next, the seeds were washed for a total of thirty minutes in a solution of 15% (volume/volume) Clorox (Clorox is an aqueous solution containing 5.25% sodium hypochlorite) and 0.1% (volume/volume) of a liquid dishwashing detergent such as Palmolive or another suitable wetting agent in distilled water. Finally, the seeds were washed five times in sterile double distilled water.

The sterilized seeds were placed embryo side up on sterile moistened Whatman No. 3 filter paper contained in a sterile Petri dish. The seeds were incubated in the covered Petri dishes in constant darkness at 25° C. for four days. The filter paper was kept moist throughout the incubation period.

#### C. Inoculation of Seedlings with *A. tumefaciens*

The two primary areas of rapidly dividing cells in the corn seedling are the root cap and the area extending from the base of the scutellar node through the mesocotyl slightly beyond the coleoptile node, the locations of which are depicted in FIGS. 1A and 1B. The mesocotyl

is the area between the scutellar node and the coleoptile node.

In the preferred practice of the invention, the wound is made in the area extending from the base of the scutellar node through and slightly beyond the coleoptile node because, included within the region, are tissues which give rise to germ line cells. In particular, found in this region is tissue which gives rise to the axillary primordia which, in turn, gives rise to tillers and to the ears (female reproductive organs). Also, located in this region is the apical meristem which gives rise to the tassel which, in turn, gives rise to the pollen (the male reproductive organs). By inoculating the seedling in this preferred area, transformation of germ line cells has been obtained.

Accordingly, a total of four wounds were made on the surface of each of the germinating corn seedlings prepared in part C of this example in an area which extends from the base of the scutellar node through and slightly beyond the coleoptile node. The wounds were made by visualizing a line which bisects this area longitudinally (the midline) as the seedling is viewed from the front as in FIG. 1B. The cuts were made perpendicular to the midline, from the midline to the outside edge of the seedling and from the front surface of the seedling, when the seedling is viewed as in FIG. 1B, through all of the tissue in the area where the cut is made. Thus, when a cut is made in the area of the scutellar node, the cut is made from the front surface through all of the tissue into the scutellum; and, when a cut is made in the mesocotyl, the cut is made from the front surface completely through all of the mesocotyl tissue. The four wounds are indicated by the numeral 1 in FIG. 1B.

The wounds were inoculated by dripping a total of 100  $\mu$ l of a  $10^9$  cells/milliliter suspension of *A. tumefaciens* strain B6 in YEB, cultured as described above in part A, onto the four wounds. As a control, some seedlings were inoculated with 0.9% NaCl (saline). After receiving the inoculum, the seedlings were placed embryo side up on a layer of Bactoagar contained in a Petri dish, at 5 seedlings per dish. The Petri dish contained 20 milliliters of sterile Bactoagar (purchased from Difco Laboratories) at a concentration of 20 grams/liter in distilled water. The covered Petri dishes were incubated at 27° C. in constant darkness for an additional 7-14 days.

#### D. Assays

1. Assay For Enzyme Activity in Seedlings: At the end of the 7-14 day incubation period, the seedlings were homogenized in a 0.1M Tris-HCl buffer, 8.0, containing 0.5M sucrose, 0.1% (weight/volume) ascorbic acid and 0.1% (weight/volume) cysteine-HCl, using a Wheaton tissue grinder, until the homogenate had a homogeneous consistency. Since the seedlings were grown in the dark, pigment formation was retarded, and cell walls remained unusually soft. Thus, the cells of the seedlings broke open easily. Next, the homogenates were spun in a Fisher Microfuge at 13,000 xg for two minutes to obtain cell-free extracts.

A portion of the cell-free extract was added to an equal volume of a reaction medium designed to detect lysopine dehydrogenase activity. This reaction medium consisted of 30 mM L-arginine, 75 mM pyruvate and 20 mM NADH dissolved in 0.2M sodium phosphate buffer, pH 7.0. The enzyme reaction was allowed to proceed at room temperature for the times indicated below.

The products of the enzyme reaction were separated electrophoretically on Whatman 3MM paper. At the start of the enzyme reaction period (time zero), a 5  $\mu$ l sample of the reaction mixture was spotted at the anodal site on the paper and dried. After 15 hours of reaction, another 5  $\mu$ l sample of the reaction mixture was spotted on the paper and dried. Finally, a 5  $\mu$ l sample of a 100  $\mu$ g/ml solution of synthetic octopine purchased from Calbiochem, Division of American Hoescht, La Jolla, Calif., and a 5  $\mu$ l sample of 100  $\mu$ g/ml solution of synthetic nopaline purchased from Sigma, St. Louis, Mo., were spotted on the paper and dried.

Electrophoresis was performed in a formic acid (90.8%)/glacial acetic acid/water (5:15:80; volume/volume) solution, pH 1.8, for 2.5 hours at 450 volts. The paper was dried and then stained by dipping it into a solution containing one part 0.02% (weight/volume) phenanthrenequinone in absolute ethanol plus one part 10% (weight/volume) NaOH in 60% (volume/volume) ethanol. After drying, the spots were visualized under a long-wave ultraviolet lamp (366 nm).

The results of the electrophoresis of the products produced by adding the cell-free extracts of B6-inoculated seedlings to lysopine dehydrogenase reaction medium are shown in FIG. 2A. In that figure, lane 1 contains a sample of the reaction mixture produced by adding a portion of the cell-free extract of ten B6-inoculated seedlings to an equal volume of the lysopine dehydrogenase reaction medium at time zero, lane 2 contains the product produced by incubating a portion of the cell-free extract of ten B6-inoculated seedlings with an equal volume of lysopine dehydrogenase reaction medium for fifteen hours, lane 3 contains the synthetic octopine, lane 4 contains the product produced by adding a portion of the cell-free extract of ten saline-inoculated seedlings to an equal volume of lysopine dehydrogenase reaction medium at time zero, lane 5 contains the product produced by incubating a portion of the cell-free extract of ten saline-inoculated seedlings with an equal volume of lysopine dehydrogenase reaction medium for fifteen hours, and lane 6 contains the synthetic nopaline.

The results shown in FIG. 2A demonstrate that octopine production is caused by cell-free extracts of B6-inoculated corn seedlings, lane 2, but that no such production occurs in the saline control, lane 5. Furthermore, the amount of octopine produced, as measured by an increase in phenanthrenequinone fluorescence, increases in proportion to the time of incubation. While no octopine can be detected at time zero, lane 1, it is clearly present after fifteen hours of incubation, lane 2. Such results are in accord with the proposition that the reaction is enzyme catalyzed and that the enzyme extracted from the B6-infected corn seedlings is lysopine dehydrogenase. Since only transformed plant tissues are known to express the opine synthase genes, these results are also in accord with the proposition that the corn seedlings have been transformed by infection with the vir+ *A. tumefaciens* strain B6.

2. Assay For Substrate Specificity: Lysopine dehydrogenase catalyzes the synthesis of octopine but not of nopaline. In FIG. 2B the results of the electrophoresis of the products produced by adding extracts of seedlings to a reaction medium containing reagents which allow for the detection of nopaline dehydrogenase enzyme activity are presented. The nopaline dehydrogenase reaction medium consists of 60 mM L-arginine, 60 mM  $\alpha$ -ketoglutarate and 16 mM NADH dissolved in

0.2M sodium phosphate buffer, pH 7.0. The  $\alpha$ -ketoglutarate can be used as a substrate only by nopaline dehydrogenase. Lanes 1-6 in FIG. 2B are the same as lanes 1-6 of FIG. 2A except that the reaction medium is the nopaline dehydrogenase reaction medium. Thus, FIG. 2B shows that cell-free extracts of seedlings inoculated with strain B6 cannot use  $\alpha$ -ketoglutarate in the condensation reaction with arginine to produce nopaline. This substrate specificity confirms the identity of the enzyme produced by seedlings transformed by strain B6 as lysopine dehydrogenase.

3. Transformation Efficiency: To address this issue, assays of single seedlings were performed, and the number of cell-free extracts of single seedlings which produced octopine were determined. The results are shown in FIG. 3A where all ten lanes contain the product produced by incubating the cell-free extracts from single B6-inoculated seedlings with lysopine dehydrogenase reaction medium for four hours. As shown in FIG. 3A, eight of the ten lanes have a spot which stains with phenanthrenequinone and co-migrates with the octopine standard. Thus, the transformation frequency for this experiment was 80%.

4. Controls: To rule out the possibility that the octopine produced was a consequence of some secondary and uninteresting event, several additional controls were done. As shown in FIG. 4A, electrophoresis of cell-free sonicates of a suspension of B6 cultured for 48 hours, as described above in part A, failed to disclose any stored octopine (lanes 1 and 2). Furthermore, when these sonicates were mixed with lysopine dehydrogenase reaction medium and incubated for four hours, no lysopine dehydrogenase activity could be detected. See FIG. 4B, lanes 1, 2 and 3 where the products of this incubation were electrophoresed. Thus, lysopine dehydrogenase activity is not found in bacterial cultures 48 hours old. Also, lysopine dehydrogenase reaction medium alone did not contain any octopine. See FIG. 4B, lane 4, where this reaction medium alone was electrophoresed. Similarly, no evidence is found for the existence of this dehydrogenase in uninfected corn seedlings. See FIG. 4C which shows an electrophoretogram in which lanes 1-5 contain the product produced by incubating cell-free extracts of uninfected single seedlings for four hours with lysopine dehydrogenase reaction medium. These results confirm that the presence of lysopine dehydrogenase activity in the cell-free extracts of B6-inoculated seedlings is due to the transformation of the seedlings and not to some secondary or uninteresting event.

## EXAMPLE II

### A. Preparation of Bacteria

A single colony of the *A. tumefaciens* strain C58 was inoculated into YEB, and the bacteria were incubated as described above in Example I, part A, for strain B6. The C58 strain is a standard wild type strain of *A. tumefaciens*. It is vir<sup>+</sup>, and it codes for the production of nopaline dehydrogenase in suitable plant hosts. It was obtained from James and Barbara Lippincott, Northwestern University, Evanston, Ill., or from Clarence Kado, University of California, Department of Plant Pathology, Davis, Calif. It has been described in Depicker et al, *Plasmid*, 3, 193 (1980) and Kao, et al, *Molec. Gen. Genet.*, 188, 425 (1982). Strain C58 is also on deposit at the ATCC and has been given accession number 33970.

### B. Transformation of Corn

Seeds of the inbred yellow lochief strain of corn were sterilized, germinated, inoculated and further incubated for 7-14 days as described above in Example I, parts B and C, except that the corn seedlings were inoculated with strain C58 rather than strain B6.

### C. Assays

1. Assay For Enzyme Activity in Seedlings: At the end of the 7-14 day incubation period, cell-free extracts of the seedlings were prepared and were assayed for enzymatic activity as described above in Example I, part D, except that the reaction medium used was the one designed to assay for nopaline dehydrogenase activity. As set forth above in Example I, part D, this reaction medium consisted of 60 mM L-arginine, 60 mM  $\alpha$ -ketoglutaric acid and 16 mM NADH dissolved in 0.2M sodium phosphate buffer, pH 7.0.

The results are shown in FIG. 2B. Lane 8 in FIG. 2B contains the product produced by mixing a portion of the cell-free extract of ten C58-inoculated seedlings with an equal volume of nopaline dehydrogenase reaction medium at time zero, and lane 7 contains the product produced by incubating a portion of the cell-free extract of ten C58-inoculated seedlings with an equal volume of nopaline dehydrogenase reaction medium for fifteen hours. Thus, FIG. 2B shows that nopaline is produced by the cell-free extracts of the C58-inoculated corn seedlings, lane 7, but that no such production was produced by the saline control, lane 5. Once again, the amount of nopaline produced, as measured by an increase in phenanthrenequinone fluorescence, increases in proportion to the time of incubation. While no nopaline can be detected in a reaction mixture at time zero, lane 8, it is clearly present after fifteen hours of incubation, lane 7, and such results are in accord with the proposition that the reaction is enzyme-catalyzed and that the enzyme extracted from C58-infected seedlings is nopaline dehydrogenase. Since only transformed plant tissues are known to express the opine synthase genes, these results are also in accord with the proposition that the corn seedlings have been transformed by infection with the vir<sup>+</sup> *A. tumefaciens* strain C58.

2. Pyronopaline Assay: The synthesis of nopaline by extracts from seedlings transformed with C58 has been confirmed by criteria other than electrophoretic mobility. When nopaline is eluted from a paper chromatogram with water, evaporated by the application of a vacuum to reduce the volume and reacted with an equal volume of hot (100° C.) 2M acetic acid for one hour, pyronopaline is formed. The conversion reaction is diagnostic for nopaline and no other opine.

In FIG. 2C, lane 1 is pyronopaline produced from synthetic nopaline by treating the synthetic nopaline with hot 2M acetic acid as described above, lane 2 is synthetic nopaline (some of which converts spontaneously to pyronopaline), lane 3 is the product produced by incubating a portion of the cell-free extract from ten C58-inoculated seedlings with an equal volume of the nopaline dehydrogenase reaction medium for fifteen hours, and lane 4 is this product treated with hot 2M acetic acid as described above. As can be seen, the product produced by incubating the cell-free extract from C58-inoculated seedlings with the nopaline dehydrogenase reaction mixture, lane 3, is totally converted to pyronopaline, lane 4, confirming that C58-inoculated seedlings produce nopaline dehydrogenase.

3. Catabolism of Nopaline: Finally, if nopaline is incubated with C58 and serves as its sole energy source, the bacteria will grow as this compound is degraded. However, B6 grown on a medium containing nopaline as the sole energy source will not break down the nopaline and will not divide since B6 lacks the specific opine oxidase which is necessary for the catabolism of nopaline.

An assay based on this principle was used to confirm the nopaline identity of the product produced by the cell-free extracts of C58-inoculated seedlings. The results of this assay are shown in FIG. 2D where lane 1 contains synthetic nopaline, lane 2 contains the product produced by incubating a portion of the cell-free extract of ten C58-inoculated seedlings with an equal volume of the nopaline dehydrogenase reaction medium for fifteen hours which has been electrophoresed, eluted with water and evaporated, as described above, in connection with the pyronopaline assay, and incubated with strain B6 for twenty-four hours, and lane 3 contains the product produced by incubating a portion of the cell-free extract of ten C58-inoculated seedlings with an equal volume of nopaline dehydrogenase reaction medium for fifteen hours which has been electrophoresed, eluted, evaporated and incubated with strain C58 for twenty-four hours. The product produced by the cell-free extracts of the C58-inoculated seedlings was consumed by strain C58, lane 3, but was not consumed by strain B6, lane 2, confirming that the product is nopaline.

4. Efficiency of Transformation: The efficiency of the transformation of inbred yellow Iochief corn using *A. tumefaciens* strain C58 was also investigated. The results are shown in FIG. 3C where all ten lanes contain the product produced by incubating cell-free extracts from single C58-inoculated seedlings with nopaline dehydrogenase reaction medium for six hours. As can be seen, 9 out of 10 seedlings were transformed. Additional single seedling assays were done using seedlings transformed with either B6 or C58. Of the 150 total single seedling assays done with either strain B6 or C58, sixty percent were transformed.

5. Controls: To rule out the possibility that the nopaline produced was a consequence of some secondary and uninteresting event, several additional controls were done. As shown in FIG. 4A, electrophoresis of cell-free sonicates of a suspension of C58 cultured for 48 hours, as described above in Example II, part A, failed to disclose any stored nopaline, lanes 3 and 4. Furthermore, when these sonicates were mixed with nopaline dehydrogenase reaction medium and incubated for four hours at 25° C., no nopaline dehydrogenase activity could be detected. See FIG. 4B, lanes 6 and 7 where the products of this incubation were electrophoresed. Thus, nopaline dehydrogenase activity was not found in bacterial cultures 48 hours old. Also, nopaline dehydrogenase reaction medium alone did not contain any nopaline. See FIG. 4B, lane 5, where this reaction medium alone was electrophoresed. Similarly, no evidence is found of this dehydrogenase in uninfected corn seedlings. See FIG. 4C which shows an electrophoretogram in which lanes 6-10 contain the product produced by incubating cell-free extracts of uninfected single seedlings for four hours with nopaline dehydrogenase reaction medium. These results confirm that the presence of nopaline dehydrogenase activity in the cell-free extracts of C58-inoculated seedlings is due to the transformation

of the seedlings and not to some secondary and uninteresting event.

### EXAMPLE III

#### A. Transformation of Corn

Strain C58 was cultured as described above in Example II, part A, and inbred yellow Iochief corn was sterilized, germinated, inoculated and incubated as described above in Example II, part B. After 7 days of incubation, the infected seedlings were planted in pots containing potting soil.

#### B. Assays

1. Assay For Enzymatic Activity in Leaves of Embryonic Origin: Three weeks after planting, three leaves of embryonic origin from three separate plants were assayed for the presence of nopaline dehydrogenase activity. The leaves chosen for the assay were the first leaves from the base of the plant which had not enlarged. The embryonic leaves are derived from differentiated structures present in the seedlings at the time of inoculation, but these differentiated structures are not located in the inoculated area.

To perform the assay, the three embryonic leaves were individually homogenized in Tris-HCl buffer, centrifuged and assayed for enzymatic activity as described above in Example II, part C, for the seedlings. The results of the electrophoresis of the product produced by incubating the cell-free extracts of the three embryonic leaves with nopaline dehydrogenase reaction medium for twelve hours are shown in FIG. 5A where lanes 1, 2 and 3 contain the products of this incubation. As can be seen from FIG. 5A, none of the cell-free extracts of the embryonic leaves contained nopaline dehydrogenase activity demonstrating that these leaves had not been transformed by the inoculation procedure.

2. Assay For Enzymatic Activity in Leaves of Meristematic Origin: Leaves derived from the meristem (all leaves besides embryonic leaves) were assayed for the presence of nopaline dehydrogenase 7 weeks after the planting of the seedlings. The meristem is tissue composed of small, rapidly dividing, undifferentiated cells which are capable of dividing to produce organs and other differentiated tissue. Meristematic tissue which differentiates into leaves is located in the inoculated area.

To perform the assay, sections were dissected out of the meristem-derived leaves, and the dissected sections from each leaf were homogenized together in Tris-HCl buffer, centrifuged and assayed for enzymatic activity as described above in Example II, part C, for seedlings. The sections of the leaves which were used for the assay are indicated by the numerals 2 and 3 in FIG. 5B. The section designated by the numeral 3 was dissected by cutting along lines 6 and 7. Line 7 is coincident with the midrib 4 of the leaf. Section 3 is located at the base of the leaf which is normally attached to the plant. The base of the leaf is the growing end of the leaf, and this area contains the newest cells on the leaf. The sections designated by the numeral 2 were dissected by cutting along line 5 which is perpendicular to the midrib 4. Sections 2, 2 are at the tip of the leaf, and they contain the oldest cells on the leaf. Each section 2 or 3 constitutes about 1/6 of the leaf's surface area.

The results of the electrophoresis of the product produced by incubating the cell-free extracts of these sec-



tions of four leaves of meristematic origin taken from four separate plants with nopaline dehydrogenase reaction medium are shown in FIG. 5C. In FIG. 5C, lane 1 contains nopaline dehydrogenase reaction medium, and lanes 2-5 contain the products produced by incubating the cell-free extracts of the leaves with nopaline dehydrogenase medium for twelve hours. As shown there, cell-free extracts of 3 out of 4 leaves produced nopaline demonstrating that they contained nopaline dehydrogenase activity. Since these leaves are derived from meristematic tissue by cell division and differentiation, these results demonstrate that the cells in the inoculated area of the seedling were able to pass on the ability to synthesize nopaline dehydrogenase to future generations of corn cells. Thus, these results demonstrate that transformation of cells in the area inoculated and of cells derived from these cells has taken place.

3. Assay For Enzymatic Activity in Pollen: Sixty days after the planting of the seedlings, samples of the pollen of two plants were individually assayed for the presence of nopaline dehydrogenase. To perform the assay, 0.5 to 1.0 milliliter of pollen containing approximately  $5-10 \times 10^5$  grains of pollen was homogenized in Tris-HCl buffer, centrifuged and assayed for enzymatic activity as described above in Example II, part C, for seedlings. The results of the electrophoresis of the products produced by incubating the cell-free extracts of the pollen from the two plants with nopaline dehydrogenase reaction medium for twelve hours are shown in FIG. 5D where lanes 2 and 3 contain the products of this incubation, and lane 1 contains nopaline dehydrogenase reaction medium. As can be seen from that figure, pollen from both of the plants contained nopaline dehydrogenase activity. Since the pollen is derived from the apical meristem by cell division and differentiation, these results, like the results for the leaves of meristematic origin above, demonstrate that transformation has taken place.

4. Assay for Enzymatic Activity in Seedlings Derived from Transformed Pollen: Pollen from the two transformed plants identified above in part B3 is used to fertilize ears on plants grown from uninfected yellow lochief corn seed. The  $F_1$  seeds produced by the fertilized plants as a result of this mating are harvested and are germinated and incubated as described above in Example I, parts B and C, except that the seedlings are not inoculated. After 7-14 days of incubation, the seedlings are assayed for enzymatic activity as described above in Example II, part C. Cell-free extracts of the seedlings are found to produce nopaline showing that this  $F_1$  generation of seedlings is transformed.

5. Assay for Enzymatic Activity in Leaves of Embryonic Origin Taken from Plants Derived From Transformed Pollen: Seeds produced by the mating described above in part B4 of this example are harvested and are germinated and incubated as described above in Example I, parts B and C, except that the seedlings are not inoculated. After 7-14 days incubation, the seedlings are planted as described above in part A of this example. Three weeks after planting, embryonic leaves are assayed for nopaline dehydrogenase activity as described above in part B1 of this example, and cell-free extracts of the embryonic leaves produce nopaline showing they are transformed.

## EXAMPLE IV

### A. Transformation of Corn

Yellow lochief corn was sterilized, germinated, inoculated and further incubated for 7-14 days as described above in Example I, Parts B and C, except that separate groups of corn seedlings were inoculated with either *A. tumefaciens* strain A348, strain JK 195 or strain 238MX.

The A348 strain carries the broad host range plasmid pTiA6NC from strain A6NC. It is vir<sup>+</sup>, and it codes for the production of lysopine dehydrogenase in suitable plant hosts. Strain A6NC and plasmid pTiA6NC are described by Sciaky et al. in *Plasmid*, 1, 238 (1977). The A348 used was obtained from Eugene Nester, University of Washington, Department of Microbiology and Immunology, Seattle, Wash. It was cultured as described above in Example I, part A, for strain B6.

Strains JK 195 and 238MX each carry a mutation in the critical vir region and are vir<sup>-</sup>. They cannot, therefore, convey the necessary portion of the Ti plasmid to their respective hosts. Consequently, plant extracts made from material inoculated with these bacteria would not be expected to produce any opine when added to the appropriate reaction medium.

The 238MX is similar in background and source to the A348 strain, but has the bacterial transposon Tn3 inserted in the vir region rendering it vir<sup>-</sup>. Strain 238MX was obtained from Eugene Nester (address given above). It was incubated as described in Example I, part A, for strain B6, and it was selected on YEB containing 100 ug/milliliter of carbenicillin.

The JK 195 strain is a vir<sup>-</sup> mutant derived from C58. It has the bacterial transposon Tn5 inserted in complementation group VI of the vir region. A detailed description of strain JK 195 may be found in Kao et al., *Mol. Gen. Genet.* 188, 425 (1982) and Lundquist et al., *Mol. Gen. Genet.*, 193, 1 (1984). The JK 195 used was obtained from Clarence Kado (address given above). It was also incubated as described in Example I, part A, and it was selected on YEB containing 50 ug/milliliter rifampicin.

### B. Assays

1. Assay For Enzyme Activity in Seedlings: As is shown in FIGS. 3B and 3D, cell-free extracts of seedlings inoculated with the 238MX strain or the JK 195 strain do not produce opines. In FIG. 3B, lanes 6-10 contain the product produced by incubating the cell-free extracts of 238MX-inoculated single seedlings with the lysopine dehydrogenase reaction medium for four hours. As can be seen, none of the seedlings was transformed since no octopine was synthesized by the cell-free extracts. In FIG. 3D, all ten lanes contain the product produced by incubating the cell-free extracts of JK 195-inoculated single seedlings with nopaline dehydrogenase reaction medium for six hours. Again, none of the seedlings was transformed since no nopaline was produced.

However, the A348 strain is competent with respect to transformation. In FIG. 3B, lanes 1-5 contain the product produced by incubating the cell-free extracts of A348-inoculated single seedlings with lysopine dehydrogenase reaction medium for four hours. As can be seen, lysopine dehydrogenase was found in the cell-free extract of one out of five single seedlings inoculated with A348 showing that the seedling was transformed.

Thus, only those vir<sup>+</sup> *A. tumefaciens* strains capable of transferring T-DNA transform corn seedlings. Those which carry mutations in the vir region and which are, therefore, transfer minus do not provoke opine synthase activity in extracts made from infected plants.

#### EXAMPLE V

A single colony of the *A. tumefaciens* strain T37 was inoculated into YEB, and the bacteria were incubated as described above in Example I, Part A, for strain B6.

The T37 strain is a standard wild type strain of *A. tumefaciens*. It is vir<sup>+</sup>, and it codes for the production of nopaline dehydrogenase in suitable plant hosts. It was originally obtained from John Kemp, University of Wisconsin, Department of Plant Pathology, Madison, Wis., and can currently be obtained from Anne C. F. Graves, University of Toledo, Dept. of Biology, Toledo, Ohio. It has been described in Turgeon et al., *PNAS*, 73, 3562 (1976) and in Sciaky et al., *Plasmid*, 1, 238 (1978).

Seeds of the inbred yellow Iochief strain of corn were sterilized, germinated, inoculated and further incubated for 7-14 days as described above in Example I, parts B and C, except that the corn seedlings were inoculated with strain T37 rather than strain B6.

At the end of the 7-14 day incubation period, the seedlings were assayed for enzyme activity as described above in Example II, part C. Using this procedure, nopaline production by the cell-free extracts of T37-inoculated corn seedlings was demonstrated showing that the seedlings were transformed.

#### EXAMPLE VI

A single colony of the *A. tumefaciens* strain C58 was inoculated into YEB, and the bacteria were incubated as described above in Example II, part A. Seeds of the inbred PA91 strain of corn were sterilized, germinated, inoculated and further incubated for 7-14 days as described above in Example I, parts B and C, for the inbred yellow Iochief strain. The PA91 strain is a standard inbred strain of corn which is commercially available. It was obtained from Jean Roberts, Eli Lilly and Co., Greenville, Ind.

At the end of the 7-14 day incubation period, the seedlings were assayed as described above in Example II, part D. As shown in FIG. 7, nopaline production by the cell-free extracts was demonstrated showing that the PA91 strain of corn had been transformed. In FIG. 6, lanes 1-10 contain the product produced by incubating the cell-free extracts of single C-58-inoculated seedlings with nopaline dehydrogenase reaction medium for six hours. As can be seen, all ten of the cell-free extracts produced nopaline showing that the seedlings were transformed.

#### EXAMPLE VII

A single colony of the *A. tumefaciens* strain B6 was inoculated into a yeast extract broth, and the bacteria were incubated as described above in Example I, part A. Seeds of the inbred yellow Iochief strain of corn were sterilized, germinated, inoculated and further incubated for 7-14 days as described above in Example I, parts B and C, except that the seeds were germinated as follows. After being sterilized, the seeds were soaked for about 12 hours in sterile distilled water. They were then incubated on sterile moistened Whatman No. 3 paper in sterile Petri dishes as described above, but they were only incubated for 1.5 to 2.0 days since soaked

seeds germinate in a shorter time than do unsoaked seeds.

At the end of the 7-14 day incubation period, the seedlings were assayed for enzyme activity as described above in Example I, part D. Using this procedure, octopine production by cell-free extracts of the infected seedlings was demonstrated showing that the seedlings were transformed.

#### EXAMPLE VIII

A single colony of the *A. tumefaciens* strain LBA 4013 was inoculated into YEB, and the bacteria were incubated as described above in Example I, part A, for strain B6. The LBA 4013 strain is a mutant strain derived from *A. tumefaciens* strain Ach5. LBA 4013 contains the wild type Ti plasmid pTiAch5 which is vir<sup>+</sup>, and LBA 4013 codes for the production of lysopine dehydrogenase in suitable plant hosts. LBA 4013 was obtained from Clegg Waldron, Eli Lilly and Co., Indianapolis, Ind. It has been described by Marton et al., in *Nature*, 277, 129 (1979).

Seeds of the inbred yellow Iochief strain of corn were sterilized, germinated, inoculated and further incubated for 7-14 days as described above in Example I, parts B and C, except that the corn seedlings were inoculated with strain LBA 4013 rather than strain B6.

At the end of the 7-14 day incubation period, the seedlings were assayed for enzyme activity as described above in Example I, part D. Using this procedure, octopine production by cell-free extracts of LBA 4013-transformed seedlings was demonstrated showing that the seedlings were transformed.

#### EXAMPLE IX

##### A. Preparation of Bacteria

A single colony of the *A. tumefaciens* strain CA19 was inoculated into YEB, and the bacteria were incubated as described above in Example I, Part A, for strain B6. The CA19 strain is derived from strain LBA 4013 and contains the pTiAch5 plasmid of LBA 4013 which is vir<sup>+</sup> as described above in Example VIII, and strain CA19 codes for the production of lysopine dehydrogenase in suitable plant hosts.

Strain CA19 also contains the micro-Ti plasmid pCEL44. Micro-plasmid pCEL44 comprises a construct consisting of the gene coding for hygromycin phosphotransferase (aphIV) inserted between the 5' promoter and associated amino terminal region-encoding sequence of an octopine synthase gene and the 3' terminator sequence of a nopaline synthase gene. This construct is assembled between T-DNA border fragments in broad-host-range vector pKT210. Micro-plasmid pCEL 44 is capable of transforming plant cells and rendering them resistant to hygromycin.

Strain CA19 is prepared as follows.

1. Culture of *Escherichia coli* RR1ΔM15/pCEL30 and Isolation of Plasmid pCEL30: Plasmid pCEL30 comprises the right-hand border sequence of the T-DNA and 5' end of the octopine synthase (ocs) gene derived from plasmid pTiA66. A linker containing a unique BglII site is fused in the 11th codon of the ocs gene. Attached to the linker are the termination and polyadenylation signals of the nopaline synthase gene of plasmid pTiC58. Attached to these sequences is a sequence which includes the left-hand border sequence of the T-DNA derived from plasmid pTiA66. A restric-

tion site and function map of plasmid pCEL30 is given in FIG. 8.

Plasmid pCEL30 can be conventionally isolated from *Escherichia coli* K12 RR1ΔM15/pCEL30. *E. coli* RR1ΔM15/pCEL30 is on deposit at the Northern Regional Research Laboratory (NRRL), Peoria, Ill. 61604, and has accession number NRRL B-15915.

The isolation is performed as follows. *E. coli* RR1ΔM15/pCEL30 is grown in 750 ml of L medium (10 g/l caesin hydrolysate, 5 g/l yeast extract, 5 g/l NaCl, 1 g/l glucose, pH 7.4) containing ampicillin at 50 mg/ml according to conventional microbiological procedures. The culture is harvested after 24 hours incubation at 37° C. with vigorous shaking.

The culture is centrifuged, and the cell pellet is resuspended in 50 ml freshly-prepared lysis buffer (50 mM Tris-HCl, pH 8, 10 mM EDTA, 9 mg/ml glucose, 2 mg/ml lysozyme). After 45 minutes incubation on ice, the suspension is mixed with 100 ml of a solution that is 0.2N NaOH and 1% SDS. The suspension is then kept on ice for a further 5 minutes. Another 90 ml of 3M sodium acetate is added, and the mixture is maintained on ice for an additional 60 minutes.

Cell debris is removed by centrifugation, and the supernatant is mixed with 500 ml ethanol. After 2 hours at -20° C., nucleic acid is pelleted by centrifugation and is resuspended in 90 ml of 10 mM Tris-HCl, pH 8, 10 mM EDTA.

The nucleic acid solution is mixed with 90 gm cesium chloride, and 0.9 ml of a solution containing 10 mg/ml of ethidium bromide. This mixture is then centrifuged at 40,000 rpm for 24 hours to purify the plasmid DNA. The plasmid DNA band is recovered and is then re-centrifuged at 40,000 rpm for 16 hours. The plasmid DNA band is again recovered and freed of cesium chloride and ethidium bromide by conventional procedures. It is next precipitated with 2 volumes of ethanol containing 90 g/l ammonium acetate. The pelleted DNA is dissolved in TE buffer (10 mM Tris-HCl, pH 8, 1 mM EDTA) at a concentration of 0.2 mg/ml.

2. Culture of *E. coli* JA221/pOW20 and Isolation of Plasmid pOW20: *E. coli* JA221/pOW20 is grown as described for *E. coli* RR1ΔM15/pCEL30 in part A1 of this example, and plasmid pOW20 is prepared as described for plasmid pCEL30 in part A1 of this example.

3. Construction of *E. coli* RR1ΔM15/pCEL40: Five μg of plasmid pCEL30 DNA are digested with 50 units of BglII restriction enzyme in a 150 μl reaction mixture of the composition recommended by the enzyme manufacturer. Restriction and other enzymes can be readily obtained from the following sources:

Bethesda Research Laboratories, Inc. Box 6010  
Rockville, Md. 20850

Boehringer Mannheim Biochemicals 7941 Castleway  
Drive P.O. Box 50816 Indianapolis, Ind. 46250

New England Bio Labs., Inc. 32 Tozer Road Beverly, Mass. 01915

Digestion is allowed to proceed for 90 minutes at 37° C.

The reaction mixture is first mixed with 8.75 μl of 0.5M Tris-HCl, pH 8, 1 mM EDTA and then with 1.25 units of calf intestinal phosphatase (which can be purchased from Boehringer Mannheim) and incubated at 37° C. for 15 minutes. The mixture is next extracted with buffered phenol, then with ether and is precipitated with 2 volumes of ethanol containing ammonium acetate. After 30 minutes at -70° C., the DNA is pelleted and redissolved in TE buffer at a concentration of 10 μg/ml.

About 20 μg of plasmid pOW20 DNA is digested with the restriction enzymes BamHI and BglII according to the enzyme manufacturer's recommended procedures to obtain the aphIV gene. The aphIV gene is an *E. coli* gene which makes plants containing the gene resistant to hygromycin.

The DNA fragments resulting from this digestion are fractionated by conventional methods of agarose gel electrophoresis and isolated by entrapment on a piece of NA-45 DEAE paper (Schleicher & Schuell, Inc., Keene, N.H. 03431) inserted into the gel during electrophoresis. DNA is eluted from the paper by spinning the paper for 5 seconds with a sufficient amount of a high salt buffer (1.0M NaCl; 0.1 mM EDTA; and 20 mM Tris, pH 8.0) to cover the paper in a microcentrifuge. The paper is incubated at 55-60° C. for 10-45 minutes with occasional swirling. The buffer is removed, and the paper washed with about 50 μl of buffer. The DNA is extracted first with phenol and then with ether and is resuspended in TE buffer at a concentration of about 25 μg/ml.

Ten ng of the phosphatase, BglII-cut plasmid pCEL30 is mixed with 50 ng of the purified ~1.3 kb BamHI-BglII fragment of plasmid pOW20 in a 15 μl ligase buffer (50 mM Tris-HCl, pH 7.6; 10 mM MgCl<sub>2</sub>; 10 mM DTT; and 1 mM ATP) containing 0.8 units of T4 DNA ligase (BRL). The mixture is incubated overnight at 15° C.

The ligation mixture is mixed with 15 μl sterile 60 mM CaCl<sub>2</sub> solution. Next, 70 μl of a suspension of competent *E. coli* RR1ΔM15 cells, which has been stored 20× concentrated in 30 mM CaCl<sub>2</sub>, 15% glycerol at -70° C., are added. After 60 minutes on ice, the transformation mixture is heat-treated at 42° C. for 2 minutes and is then incubated with 0.5 ml L medium for 90 minutes at 37° C.

Samples of the mixture are spread on L medium containing ampicillin at 50 mg/l and solidified with agar at 15° C. These samples are then incubated overnight at 37° C. to permit growth of colonies from transformed cells.

Colonies resulting from the transformation are inoculated into 5 ml L medium containing ampicillin at 50 mg/ml and grown overnight at 37° C. Plasmid DNA is prepared from 1 ml samples of these cultures by the procedure of Holmes & Quigley, *Analytical Biochemistry*, 114, 193 (1981) and is redissolved in 50 μl of TE buffer.

4. Construction of Micro-Ti Plasmid pCEL44: Since plasmid pCEL40 is not capable of replication in *Agrobacterium*, the micro T-DNA of plasmid pCEL40 was first transferred, as an EcoRI fragment, into broad-host-range vector pKT210. This broad-host-range vector is available from Plasmid Reference Center, Stanford University, Palo Alto, Calif. 94305.

Five μg of plasmid pKT210 are digested with 50 units of EcoRI restriction enzyme in a 150 μl reaction of a composition recommended by the enzyme manufacturer. After 90 minutes at 37° C., the reaction is treated with calf intestinal phosphatase as described above in part A3 of this example and is dissolved in TE buffer at a concentration of 10 μg/ml.

Fifteen μl of a preparation of plasmid pCEL40 DNA, grown as described above in part A3 of this example, are digested with 10 units of EcoRI restriction enzyme in a 20 μl reaction at 37° C. for 90 minutes and are then extracted with phenol, followed by extraction with ether. The digested DNA is precipitated with 2 vol-

umes of ethanol containing ammonium acetate at  $-20^{\circ}\text{C}$ . and is redissolved in 20  $\mu\text{l}$  TE buffer.

Ten ng of phosphatased, EcoRI-cut pKT210 are ligated with 5  $\mu\text{l}$  of EcoRI-cut pCEL40 as described above in part A3 of this example, and transformed into *E. coli* RR1 $\Delta$ M15 as described above in part A3 of this example.

Transformed cells containing pCEL44 are selected by their ability to grow on solidified L medium containing chloramphenicol at 10 mg/l. A restriction site and function map of pCEL44 is provided in FIG. 9.

5. Conjugation of pCEL44 Into *A. tumefaciens* LBA4013 to form Strain CA19: *E. coli* K12 RR1 $\Delta$ M15/pCEL44 and *E. coli* pRK2013 are grown overnight at  $37^{\circ}\text{C}$ . on solidified L medium. *A. tumefaciens* LBA4013 is grown for 2 days at  $28^{\circ}\text{C}$ . on solidified L medium.

One loop of *E. coli* K12 RR1 $\Delta$ M15/pCEL44, one loop of *E. coli* pRK2013 and one loop of *A. tumefaciens* LBA 4013 are mixed in 1 ml of 30 mM magnesium sulfate solution. Next, a drop of the mixture is placed on solidified TY medium (5 g/l caesin hydrolysate, 5 g/l yeast extract, 15 g/l agar) and incubated at  $28^{\circ}\text{C}$ . overnight.

The bacterial growth is resuspended in 3 ml of 10 mM magnesium sulfate solution and 0.1 ml samples of serial dilutions are spread on solidified TY medium containing 100 mg/l nalidixic acid and 2 mg/l chloramphenicol and incubated at  $28^{\circ}\text{C}$ .

Transconjugants give rise to individual colonies after 2 to 4 days growth. These are inoculated singly into 25 ml liquid TY medium containing 100 mg/l nalidixic acid and 2 mg/l chloramphenicol and incubated at  $28^{\circ}\text{C}$ . with shaking for another 2 days. The plasmid content of the transconjugants is then checked by the method of Casse et al. (*Journal of General Microbiology* 113:229-242; 1979), and strain CA19 containing the wild type pTiAch5 plasmid and the pCEL44 plasmid is isolated.

The CA19 used to practice the method of the present invention was obtained from Clegg Waldron, Eli Lilly and Co., Indianapolis, Ind. The preparation of strain CA19 has also been described in Waldron et al., *Plant Molec. Biol.*, 5, 103 (1985) which is incorporated herein by reference.

#### B. Transformation of Corn

Seeds of the inbred yellow Iochief strain of corn were sterilized, germinated, inoculated and further incubated for 7-14 days as described above in Example 1, parts B and C, except that the corn seedlings were inoculated with strain CA19 rather than strain B6.

#### C. Assays

1. Assay of Seedlings for Lysopine Dehydrogenase Activity: At the end of the 7-14 day incubation period, the seedlings were assayed for enzyme activity as described above in Example 1, part D. The results are shown in FIG. 6 where lanes 1-10 contain the product produced by incubating cell-free extracts of single CA19-inoculated seedlings with lysopine dehydrogenase for four hours. As can be seen there, octopine production in 9 out of 10 cell-free extracts of single CA19-inoculated corn seedlings was demonstrated showing that the seedlings contained lysopine dehydrogenase and were transformed.

2. Assay of Seedlings for Hygromycin Resistance: Some seedlings which were inoculated with strain

CA19 were incubated only for 3-4 days after inoculation, at which time they were assayed for resistance to hygromycin as follows. The seedlings were dissected away from the endosperm and scutellum (see FIGS. 1A and 1B) and cut into approximately 3 mm cross sections. The cross sections were cultured for three weeks on Duncan's medium (described by Duncan et al. in *Planta*, 165, 322 (1985)) supplemented with 200  $\mu\text{g}/\text{ml}$  each carbenicillin (Sigma) and vancomycin (Lilly) or on BN4 medium (Murashige and Skoog major and minor salts (described in *Physiol. Plant*, 15 473 (1962)), 4 mg/l 2,4-dichlorophenoxy acetic acid as auxin, 9 g/l Difco Bactoagar and 20 g/l sucrose) supplemented with 200  $\mu\text{g}/\text{ml}$  each carbenicillin and vancomycin, in the dark, at  $25^{\circ}$  to  $27^{\circ}\text{C}$ ., followed by another three-week passage on those antibiotics.

To test the response to hygromycin of the tissue cultures derived from the corn seedlings, either whole cross sections plus the tissue which has grown up from the cross sections or 100 mg callus samples are placed onto about 50 ml of Duncan's medium or of BN4 medium supplemented with the aforementioned concentrations of vancomycin and carbenicillin and containing about 125  $\mu\text{g}/\text{ml}$  of hygromycin B (Lilly) contained in Falcon 1005 Petri dishes. This test is read after three weeks of incubation in the dark at  $27^{\circ}\text{C}$ ., by visually checking for growth. Cultures that are growing are light in color and show an increase in size. Using this test, positive growth phenotypes are recovered from cultures derived from CA19-inoculated seedlings showing that the seedlings are transformed by the heterologous hygromycin gene.

#### EXAMPLE X

A. Preparation of an *A. tumefaciens* Strain Carrying the Gene Conferring Resistant to the Herbicide Glyphosate

1. Culture of *E. coli* RR1 $\Delta$ M15/pCEL30 and Isolation of Plasmid pCEL30: *E. coli* RR1 $\Delta$ M15/pCEL30 is grown as described above in Example IX, part A1, and plasmid pCEL30 is isolated as described in Example IX, part A1.

2. BqIII Digestion of Plasmid pCEL30 and Treatment With Calf Intestinal Phosphatase: Five  $\mu\text{g}$  of plasmid pCEL30 DNA are digested and treated with calf intestinal phosphatase as described above in Example IX, part A3.

3. Isolation of Glyphosate-Resistant EPSP Synthase Gene: A gene coding for a glyphosate-resistant 5-enolpyruvylshikimate 3-phosphate synthase gene (GRESPS gene) is isolated as described by Comai et al. in *Nature*, 317, 714 (1985) and by Stalker et al., *J. Biol. Chem.*, 260, 4724 (1985) which are incorporated herein by reference. In the final steps of this procedure, the GRESPS gene, in a BamHI fragment cut from plasmid pPMG34, is cloned into plasmid pUC7 to give plasmid pPMG38. The GRESPS gene is then excised from plasmid pPMG38 as an EcoRI fragment. The EcoRI fragment is then modified using a suitable commercially available molecular linker so that it is able to ligate with the unique BqIII site on the BqIII-digested pCEL30 plasmid and so that the EcoRI site is removed.

The herbicide glyphosate (N-phosphonomethylglycine) is a widely used broad-spectrum herbicide that kills both weed and crop species. It inhibits a metabolic step in the biosynthesis of aromatic compounds, and the cellular target of glyphosate is 5-enolpyruvyl shikimate 3-phosphate synthase (EPSP synthase) which

catalyzes the formation of 5-enolpyruvylshikimate 3-phosphate from phosphoenolpyruvate and shikimate, and inhibition of this step of the shikimate pathway eventually leads to cellular death. The GRESPS gene is a mutant allele of the *aroA* locus of *Salmonella typhimurium* which encodes a EPSP synthase in which the substitution of a serine for proline causes a decreased affinity of the enzyme for glyphosate.

4. Ligation: Ten ng of the phosphatased, BglII-cut plasmid pCEL30, as prepared above in Example IX, part A3, are mixed with 50 ng of the GRESPS gene (including the attached linker) in a 15  $\mu$ l ligase buffer containing 0.8 units of T4 DNA ligase. The mixture is incubated overnight at 15° C. The ligation mixture is used to transform competent *E. coli* RR1 $\Delta$ M15 cells as described in Example IX, part A3.

5. Construction of Micro-Ti Plasmid Carrying the GRESPS Gene: After selection of transformed cells produced as described in part A4 of this example on L medium containing ampicillin as described above in Example IX, part A3, the plasmids from the transformed *E. coli* RR1 $\Delta$ M15 cells are transferred as described above in Example IX, part A4, as an EcoRI fragment, into broad-host-range vector pKT210.

6. Conjugation Into *A. tumefaciens* LBA 4013: The plasmid carrying the GRESPS gene on broad-host-range vector pKT210 is transferred into *A. tumefaciens* strain LBA 4013 by conjugation as described above in Example IX, part A5. The resulting transconjugants are selected as described above in Example IX, part A5, and a new strain of *A. tumefaciens* carrying the GRESPS gene which is herein referred to as strain LBA 4013/GRESPS, is isolated.

#### B. Transformation of Corn

Seeds of the inbred yellow Iochief strain of corn are sterilized, germinated, inoculated and further incubated as described above in Example I, parts B and C, except that the corn seedlings were inoculated with strain LBA 4013/GRESPS rather than strain B6. After 7 days of incubation, the infected seedlings are planted in pots in potting soil.

#### C. Assays

1. Assay For Enzymatic Activity in Pollen: Sixty days after the planting of the seedlings, samples of the pollen of five plants are individually assayed for the presence of lysopine dehydrogenase as described above in Example III, part B. The results of the electrophoresis of the products produced by incubating the cell-free extracts of the pollen of the five plants show that three out of five of the cell-free extracts of the pollen produce octopine showing that the pollen is transformed.

2. Assay for EPSP Synthase Activity: Leaves derived from the meristem are assayed seven weeks after planting of the seedlings for EPSP synthase activity according to the method of Boocock and Coggins, *FEBS Letters*, 154, 127 (1983) which is incorporated herein by reference. Five leaves from five separate plants are assayed, and three are found to contain EPSP synthase activity showing that transformation had occurred.

3. Assay For Resistance To Glyphosate: The three plants found to contain EPSP synthase activity in their leaves as a result of the assay in part C2 of this example are sprayed with the equivalent of 0.5 kg/hectare of the isopropylamine salt of glyphosate. All three plants show considerable tolerance to glyphosate as compared to controls.

None of the foregoing description of the preferred embodiments is intended in any way to limit the scope of the invention which is set forth in the following claims. Those skilled in the art will recognize that many modifications, variations and adaptations are possible.

#### EXAMPLE XI

##### A. Transformation of Corn

Strain CA19 was cultured as described in Example IX, part A, and yellow Iochief corn was sterilized, germinated, inoculated with CA19 and inoculated as described in Example I, parts B and C. Seedlings were also inoculated with *A. tumefaciens* strain CA17 which is identical to strain CA19, except that the gene that codes for hygromycin resistance is inserted in the antisense direction. Finally, seedlings were also inoculated with YEB alone. After seven days of incubation, all of the inoculated seedlings were planted.

##### B. Assays

1. Assay For Enzymatic Activity in the Upper Leaves of Transformed Plants: As the plants reached sexual maturity, their upper leaves were assayed for the presence of lysopine dehydrogenase activity as described in Example III, part B2, except that lysopine dehydrogenase reaction medium was used. Leaves of five plants derived from seedlings inoculated with CA19, of two plants derived from seedlings inoculated with CA17 and of three plants inoculated with YEB were assayed.

The results are shown in FIG. 14. In FIG. 14, the lanes contain the following materials:

Lane	Plant Code No.	CONTENTS OF LANE		Comments
		Proculum Used on Seedling	Plant Part Sampled	
1	—	—	—	Synthetic octopine standard
2	S-1	YEB	Flag leaf	—
3	19-6	CA19	Leaf below flag leaf	—
4	19-6	CA19	Tiller	—
5	17-4	CA17	Ear shoot	—
6	19-3	CA19	Ear shoot	—
7	19-3	CA19	Leaf below flag leaf	—
8	S-5	YEB	Leaf below flag leaf	—
9	17-3	CA17	Leaf below flag leaf	—
10	17-3	CA17	Second leaf below flag leaf	—
11	19-4	CA19	Leaf below flag leaf	—
12	S-8	YEB	Leaf below flag leaf	—
13	19-5	CA19	Flag leaf	—
14	19-5	CA19	Leaf below flag leaf	—
15	19-2	CA19	Flag leaf	—
16	19-2	CA19	Leaf below flag leaf	—

As shown in FIG. 14, none of the extracts of leaves derived from YEB-inoculated seedlings produced octopine, whereas 8 out of 12 extracts of leaves derived from CA19-inoculated and CA-17-inoculated seedlings produced octopine.

2. Assay For Bacteria In Upper Leaves of Transformed Plants: Aliquots of the extracts of the leaves used in the lysopine dehydrogenase assay described in part B2 of this example were also plated to determine if any bacteria were present in these extracts. Any bacterial colonies growing up as a result of these platings were transferred to a lactose-containing medium diagnostic for *Agrobacterium*. None of the bacteria that grew up as a result of the original platings (either in extracts of leaves derived from YEB-inoculated seedlings, from CA19-inoculated seedlings, or CA17-inoculated seedlings) oxidized lactose to lactic acid showing that none of them were *Agrobacterium* of the type used to inoculate the seedlings.

3. Assay For Enzymatic Activity in Leaves of F<sub>1</sub> Plants: Using the results of the above assay, plants 19-5 and 19-3 were chosen for further study. Plant 19-5 was chosen because extracts of both of the two top leaves were positive for lysopine dehydrogenase activity indicating that this plant might have a transformed sector extending into the tassel. Plant 19-3 was chosen because extracts of its ear shoot and leaf below the flag leaf were positive for lysopine dehydrogenase activity indicating that this plant might have a transformed sector extending into the ear and a transformed sector extending into the tassel.

These two plants were self pollinated. Then, 26-27 days post pollination, the immature progeny ears of plants 19-3 and 19-5 were removed from the plants and surface sterilized. The late maturity embryos of the ears from these plants were excised and planted on half-strength Murashige and Skoog medium. The embryos were allowed to germinate sterily in the light for 8-10 days, at which time they were planted in soil.

The meristem-derived leaves of the F<sub>1</sub> seedlings which survived transplanting to soil were assayed for lysopine dehydrogenase activity as described in Example III, part B2, except that lysopine dehydrogenase reaction medium was used. The intensity of the staining of the spots on the electrophoretogram that comigrated with octopine was rated. The results are presented in Table 1. As shown there, some of the leaves of the seedlings produced octopine, showing the sexual transmission of this trait to the F<sub>1</sub> generation.

Rating	Leaves From Plants Produced By Embryos Taken From Plant 19-3	Leaves From Plants Produced By Embryos Taken From Plant 19-5
Dead (no test)	85	9
Negative	71 (25 subsequently died)	76 (7 subsequently died)
Weak Positive	5 (3 subsequently died)	17 (0 subsequently died)
Positive	34 (8 subsequently died)	27 (3 subsequently died)
Strong Positive	3 (0 subsequently died)	3 (0 subsequently died)

## TRANSFORMATION OF OTHER SPECIES OF GRAMINEAE

### EXAMPLE XII

A single colony of the *A. tumefaciens* strain B6 was inoculated into a yeast extract broth, and the bacteria were incubated as described above in Example I, part A. Seeds of rye were sterilized, germinated, inoculated and further incubated for 7-14 days as described above

in Example I, parts B and C. The seedlings were inoculated in the apical meristem, an area of rapidly dividing cells that gives rise to the germ line cells.

At the end of the 7-14-day incubation period, the seedlings were assayed for enzyme activity as described above in Example I, part D. The results of the electrophoresis of the products produced by adding the cell-free extracts of the B6-inoculated rye seedlings to lysopine dehydrogenase reaction medium are shown in FIG. 10. In that figure, lane 1 contains the synthetic octopine standard and lanes 2-6 contain the product produced by incubating the cell-free extract of single B6-inoculated rye seedlings with lysopine dehydrogenase reaction medium. The results shown in FIG. 10 demonstrate that octopine production is caused by the cell-free extracts of three out of four B-6-inoculated rye seedlings tested (lanes 2, 3, 4 and 6). These results show that the rye seedlings have been transformed by infection with the vir<sup>+</sup> *A. tumefaciens* strain B6.

### EXAMPLE XIII

A single colony of the *A. tumefaciens* strain B6 was inoculated into a yeast extract broth, and the bacteria were incubated as described above in Example I, part A. Barley seeds were sterilized, germinated, inoculated and further incubated for 7-14 days as described above in Example I, parts B and C. The seedlings were inoculated in the apical meristem, an area of rapidly dividing cells that gives rise to the germ line cells.

At the end of the 7-14-day incubation period, the seedlings were assayed for enzyme activity as described above in Example I, part D. The results of the electrophoresis of the products produced by adding the cell-free extracts of B6-inoculated barley seedlings to lysopine dehydrogenase reaction medium are shown in FIG. 11. In that figure, lane 1 contains the synthetic octopine standard and lanes 2-6 contain the product produced by incubating the cell-free extract of single B6-inoculated barley seedlings with lysopine dehydrogenase reaction medium. The results shown in FIG. 11 demonstrate that octopine production is caused by five out of five cell-free extracts of B6-inoculated barley seedlings tested and show that the seedlings have been transformed.

### EXAMPLE XIV

A single colony of the *A. tumefaciens* strain C58 was inoculated into YEB, and the bacteria were incubated as described above in Example II, part A. Oat seeds were sterilized, germinated, inoculated and further incubated for 7-14 days as described above in Example I, parts B and C. The seedlings were inoculated in the apical meristem, an area of rapidly dividing cells that gives rise to the germ line cells.

At the end of the 7-14-day incubation period, the seedlings were assayed as described above in Example II, part D. The results are shown in FIG. 12. In FIG. 12, lane 1 contains nopaline dehydrogenase reaction medium alone, lane 2 contains the nopaline standard, lanes 3-11 contain the product produced by incubating the cell-free extracts of single C58-inoculated oat seedlings with nopaline dehydrogenase reaction medium and lane 12 contains synthetic octopine. As can be seen, six out of nine of the cell-free extracts of single C53-inoculated oat seedlings produced nopaline showing that the seedlings were transformed.

## EXAMPLE XV

A single colony of the *A. tumefaciens* strain C58 was inoculated into YEB, and the bacteria were incubated as described above in Example II, part A. Wheat seeds were sterilized, germinated, inoculated and further incubated for 7-14 days as described above in Example I, parts B and C. The seedlings were inoculated in the apical meristem, an area of rapidly dividing cells that gives rise to the germ line cells.

At the end of the 7-14-day incubation period, the seedlings were assayed as described above in Example II, part D. The results are shown in FIG. 13. In FIG. 13, lane 1 contains the nopaline standard, lanes 2-6 contain the product produced by incubating the cell-free extracts of single C58-inoculated wheat seedlings with nopaline dehydrogenase reaction medium, lane 7 contains synthetic octopine, lane 8 contains a mixture of synthetic octopine and lysopine dehydrogenase reaction medium and lanes 9-15 contain the product produced by incubating the cell-free extracts of single C58-inoculated wheat seedlings with lysopine dehydrogenase reaction medium. As can be seen, three out of five of the cell-free extracts of the C58-inoculated wheat seedlings produced nopaline when incubated with nopaline dehydrogenase reaction medium, show-

ing that the seedlings were transformed. None of the cell-free extracts produced octopine when incubated with lysopine dehydrogenase reaction medium.

We claim:

1. A method of producing transformed Gramineae, said method comprising:

making a wound in a graminaceous seedling with newly emerged radicle and stem, said wound being made in an area of the seedling containing rapidly dividing cells, wherein said area extends from the base of the scutellar node to slightly beyond the coleoptile node; and

inoculating the wound with vir+ *Agrobacterium tumefaciens*.

2. The method of claim 1 wherein about four wounds are made in the seedling, and a total of about  $10^8$  *Agrobacterium tumefaciens* cells are used to inoculate the wounds.

3. The method of claim 1 wherein the vir+ *Agrobacterium tumefaciens* contains a vector comprising genetically-engineered T-DNA.

4. The method of claim 1 or 3 wherein the Gramineae is selected from the group consisting of corn, wheat, rye, barley and oats.

\* \* \* \* \*

30

35

40

45

50

55

60

65

a partially inbred initial population, the results were similar to those observed (Table 1).

With the asymmetrical overdominance model, populations at equilibrium under high selfing do not maintain variation, but will fix the allele least disadvantageous in homozygotes though an increase will occur if the set of lines has not reached equilibrium, so that some still remain with the lower fitness allele. One would thus expect no increase in fitness on intercrossing the inbred lines after the population reaches equilibrium. During inbreeding, the fitness values behaved similarly to those in the mutational model, with a slight decrease, rather than increase, in fitness (Fig. 2c; Table 1). But the intercross mean fitness, although higher than the inbred fitness, was never higher than the fitness of the initial population, unlike the mutational load model and the observed data (Table 1).

These results support the partial dominance (mutational) hypothesis<sup>2,7</sup> for genetic load, and suggest that the mutation rate per genome in *E. paniculata* must be very high (of the order of one per generation). This is consistent with most other available data<sup>15,16</sup>. The failure of fitness to recover during inbreeding, and intercrossed fitness levels higher than those of the original outbred strains, are also seen in maize<sup>17</sup> and mice<sup>9</sup> (Table 1). This suggests that purging of partially recessive mutations accounts for an important component of heterosis. When the fitness values under inbreeding are plotted against the inbreeding coefficient, the partial dominance model yields an increase with high *F* values, due to purging, but the overdominance model predicts a monotonic decline. The partial dominance model can thus also explain the observed patterns in inbreeding experiments with conifers<sup>18</sup>.

The occurrence of natural selection during the inbreeding experiment, consistent with purging, was indicated by excess heterozygosity of five (presumably neutral) allozyme marker loci scored in the experimental plants, each generation of inbreeding (data not shown). Under inbreeding, heterozygotes tend to be produced by outcrosses. Surviving adults in the progeny generation therefore tend to be heterozygotes<sup>19,20</sup>, and excess heterozygosity compared with the neutral expectation is frequently found in inbreeding plants<sup>21</sup>.

These results are also relevant to populations undergoing disturbance by humans. Inbreeding of formerly outbred populations occurs in many zoo populations, as well as domesticated populations and populations subjected to severe reductions in size as a result of human alterations of the environment. Extreme inbreeding has been recommended to purge genetic load and force the adaptation of endangered populations to the inbreeding regime they will experience under human management<sup>22</sup>. This assumes that lowered fitness inevitably caused by increased homozygosity during this process will not be too severe or prolonged. The validity of this assumption depends on the severity and dominance of the mutant alleles in the population being inbred<sup>23,5</sup>. Our results indicate that purging by the most severe inbreeding, self-fertilization, could decrease fitness considerably, with little recovery under inbreeding, and that fitness is restored only when the inbred lines are intercrossed. □

Received 10 June; accepted 21 June 1991.

- East, E. M. & Jones, D. F. *Inbreeding and Outbreeding: Their Genetic and Sociological Significance* 147 (Lippincott, Philadelphia, 1919).
- Crow, J. F. *Genetics* 33, 477-487 (1948).
- Wright, S. *Evolution and the Genetics of Populations* Vol. 3 (Univ. of Chicago Press, 1977).
- Lande, R. & Schermse, D. W. *Evolution* 39, 24-40 (1985).
- Charlesworth, D., Morgan, M. T. & Charlesworth, B. *Evolution* 44, 1469-1489 (1990).
- Crow, J. F. in *Mathematical Models in Population Genetics* (eds Kojima, K.-I.) 128-177 (Springer-Verlag, Berlin, 1970).
- Charlesworth, D. & Charlesworth, B. A. *Rev. ecol. Syst.* 18, 237-268 (1987).
- Ziehe, M. & Roberts, J. H. *Genetics* 121, 861-868 (1989).
- Falconer, D. S. *Genet. Res. Camb.* 17, 215-235 (1972).
- Hollingsworth, M. J. & Maynard-Smith, J. *J. Genet.* 53, 295-314 (1955).
- Bryant, E. H., Meffert, L. M. & McCommas, S. A. *Amer. Nat.* 136, 542-549 (1990).
- Barrett, S. C. H. & Husband, B. C. *Pl. Spec. Biol.* 5, 41-55 (1990).
- Glover, D. E. & Barrett, S. C. H. *Heredity* 59, 7-17 (1987).
- Charlesworth, D. & Charlesworth, B. *Evolution* 44, 870-888 (1990).
- Kondrashov, A. *Nature* 336, 435-440 (1988).

- Charlesworth, B., Charlesworth, D. & Morgan, M. T. *Nature* 347, 380-382 (1990).
- Shull, G. H. in *Heterosis* (ed. Gowen, J. W.) 14-48 (Iowa University Press, Ames, 1950).
- Griffin, A. R. & Lindgren, D. *Theor. appl. Genet.* 71, 334-343 (1985).
- Ohta, T. & Charlesworth, B. *Genet. Res.* 61, 1-11 (1990).
- Sli
- Br
- Te
- Wi
- Ko

## Claim 47 Aequorin

ACKNO  
Science  
D.C. We  
on the r.

itural  
nt to  
ents

## Transgenic plant aequorin reports the effects of touch and cold-shock and elicitors on cytoplasmic calcium

Marc R. Knight\*, Anthony K. Campbell†, Steven M. Smith\* & Anthony J. Trewavas\*

\* Institute of Cell and Molecular Biology, University of Edinburgh, Edinburgh EH9 3JH, UK

† Department of Medical Biochemistry, University of Wales College of Medicine, Cardiff CF4 4XN, UK

**METHODS** for measuring plant cytoplasmic calcium using microelectrodes or microinjected fluorescent dyes are associated with extensive technical problems, so measurements have been limited to single or small groups of cells in tissue strips or protoplasts<sup>1,2</sup>. Aequorin is a calcium-sensitive luminescent protein<sup>3</sup> from the coelenterate *Aequorea victoria* (*A. forskalea*) which is formed from apoaquorin, a polypeptide of relative molecular mass ~22,000, and coelenterazine, a hydrophobic luminophore<sup>4</sup>. Microinjected aequorin has been widely used for intracellular calcium measurement in animal cells<sup>4</sup>, but its use in plants has been limited to exceptionally large cells<sup>5</sup>. We show here that aequorin can be reconstituted in transformed plants and that it reports calcium changes induced by touch, cold-shock and fungal elicitors. Reconstituted aequorin is cytoplasmic and nonperturbing; measurements can be made on whole plants and a calcium indicator can be constituted in every viable cell. Now that apoaquorin can be targeted to specific organelles, cells and tissues, with the range of coelenterazines with differing calcium sensitivities and properties available<sup>6</sup>, this new method could be valuable for determining the role of calcium in intracellular signalling processes in plants.

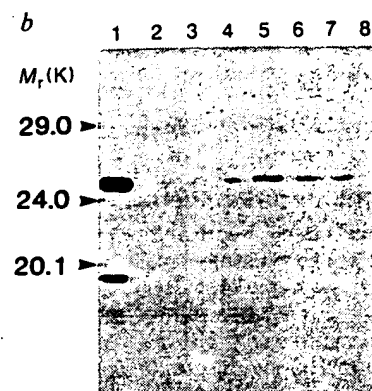
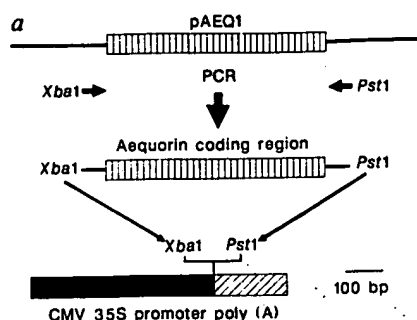
The apoaquorin-coding region from complementary DNA clone pAEQ1 (ref. 7) was fused to the cauliflower mosaic virus (CMV) 35S promoter<sup>8</sup> and transferred to *Nicotiana glauca* using the *Agrobacterium tumefaciens* pBIN19 binary vector system<sup>9</sup> to provide constitutive expression (Fig. 1a). Expression was detected by western blots, using antibodies raised against aequorin purified from *Aequorea*<sup>10</sup>, in all plants transformed with the 35S-apoaquorin chimaeric gene (Fig. 1b). *F*<sub>1</sub> progeny from the transformant expressing the highest levels of apoaquorin (MAQ2.4; Fig. 1b) were selfed and homozygous *F*<sub>2</sub> progeny were used for subsequent experiments.

The time course for aequorin reconstitution from apoaquorin in whole seedlings, in seedling homogenates and from apoaquorin purified from *Aequorea*, all show similar kinetics (Fig. 2). Seedlings incubated in coelenterazine were transferred back to agar and continued to grow vigorously and respond tropically to light and gravity. No signs of toxicity were detected and growth rates of coelenterazine-treated and untreated seedlings were similar. The ratio of reconstituted aequorin in cotyledons, shoots and roots was 1:1:0.25 on a fresh-weight basis.



FIG. 1. Construction of chimaeric gene used to express apoaequorin in *Nicotiana*. *b*, Expression of apoaequorin in *Nicotiana*. Western blot analysis of total protein from the  $F_1$  progeny of separate transformants. Lane 1, total protein from *E. coli* expressing apoaequorin<sup>19</sup>; lanes 2–7, separate *Nicotiana* 35S-apoaequorin transformants, MAQ3.3, MAQ3.2, MAQ2.13, MAQ2.4, MAQ2.3 and MAQ2.1; lane 8, wild-type untransformed *Nicotiana*.

**METHODS.** The apoaequorin coding region was amplified by polymerase chain reaction (PCR) from cDNA clone pAEQ1 (ref. 7) (the kind gift of Milton Cormier, University of Georgia), before construction of this chimaeric gene, to remove homopolymeric dC–dG tails present at either end of the cDNA insert. This counteracted inhibition of expression detected in constructs containing these tails (not shown). The *Xba*I–*Pst*I fragment was transferred to pDH51 (ref. 20) containing the CMV 35S promoter and terminator and the whole chimaeric gene was transferred to pBIN19 as an *Eco*R1 insert. All recombinant DNA manipulations and *Nicotiana* genetic transformations followed standard procedures<sup>21,22</sup>. Total protein was prepared from 1 mg of 8-day transformant seedling tissue, electrophoresed on a 10% SDS-polyacrylamide gel and transferred to a nitrocellulose filter by electroblotting. After overnight incubation in a 1:1,000 dilution of mouse primary antibody



to aequorin isolated from *Aequorea*, the filter was incubated in a 1:5,000 dilution of alkaline phosphatase conjugated anti-mouse IgG for 2 h. Colour was developed by incubation in 100  $\mu$ g ml<sup>-1</sup> nitroblue tetrazolium, 60  $\mu$ g ml<sup>-1</sup> BCIP (5-bromo-4-chloro-3-indolyl phosphate), 4 mM MgCl<sub>2</sub>, 100 mM Tris, pH 9.5. Protein electrophoresis and western blot analysis were according to standard procedures<sup>23</sup>.

The cellular locale of the apoaequorin was estimated by centrifuging homogenates of transgenic seedlings at 116,000g and estimating distribution with coelenterazine by reconstitution. Over 97% of apoaequorin was contained in the soluble fraction.

To investigate the effect of touch, seedlings were placed in a luminometer and the cotyledons touched with a fine wire or plastic rod at intervals of 1 min (Fig. 3a). With each touch an intracellular calcium concentration ( $[Ca^{2+}]_i$ ) spike was observed. When repeatedly touched on the cotyledons at about 1-s intervals, a maximal response was elicited with about four touches (data not shown). Slight movement of the cotyledons on touching seems essential to this response. Touch responses are often thought to be limited to tendrils and sensitive plants such as *Mimosa pudica* and the Venus fly trap (*Dionaea muscipula*). Recent data<sup>11</sup>, however, show that a variety of touch stimuli substantially reduces growth of *Arabidopsis* and other plants. In this case the involvement of  $[Ca^{2+}]_i$  in transducing touch signals was surmised from increased calmodulin gene expression. The data in Fig. 3a support this hypothesis much more directly.

The effect of imposing temperature fluctuations on seedlings from an ambient 20°C to temperatures from 0°C to 50°C is shown in Fig. 3b. Dramatic increases in  $[Ca^{2+}]_i$  accompany the transition to 0°C or 5°C. Such plants can be rewarmed and on recooling exhibit a further  $[Ca^{2+}]_i$  spike. The spike is higher if 10 mM  $Ca^{2+}$  is included in the chilling medium but has no effect at higher temperatures. Only irrigation with  $Ca^{2+}$  at 100 mM has any effect at 20°C, causing repetitive spiking. No large increases due to heat-shock temperatures (40°C+) were observed. Many agronomically important plants are chill-sensitive and it has been suggested that chilling injury results from a failure in  $[Ca^{2+}]_i$  homeostasis<sup>12</sup>. This method of  $[Ca^{2+}]_i$  measurement will permit a proper test of this hypothesis in chill-sensitive plants. Temperate plants use fluctuating temperature regimes as signals modifying dormancy, reproduction and inducing temperature adaptation. A role for  $[Ca^{2+}]_i$  in transducing these signals is now clearly indicated.

A variety of fungal elicitors which initiate host-plant defence responses of hydrolytic enzyme production<sup>13</sup>, phytoalexin production and lignification<sup>14</sup> have been reported in many species including tobacco<sup>15</sup>. Elicitor preparations from yeast and *Gliocladium deliquescens* have been shown to stimulate phytoalexin production in plant cells<sup>16,17</sup>. These elicitor preparations also

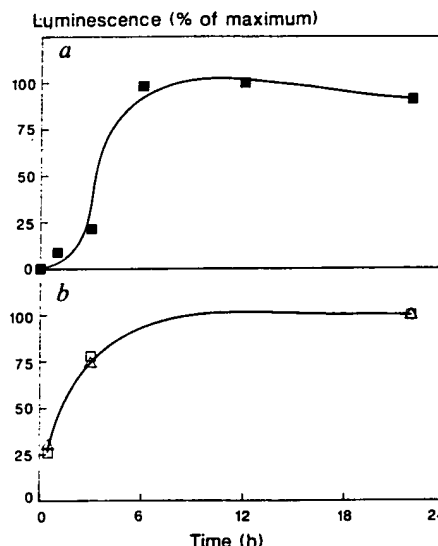


FIG. 2. Reconstitution of aequorin from apoaequorin in whole tobacco seedlings, tobacco homogenates and from purified *Aequorea* apoaequorin. *a*, Reconstitution time course of aequorin in whole *Nicotiana* seedlings (■). *b*, Reconstitution time course *in vitro* of apoaequorin isolated from *Aequorea* (Δ) and *Nicotiana* (□). Reconstitution was estimated by measuring the amount of calcium-dependent luminescence produced from the reconstituted aequorin at each time point.

**METHODS.** Apoaequorin was produced from *Aequorea* aequorin as previously described<sup>24</sup>. Reconstitutions of apoaequorin in *Nicotiana* homogenates and from *Aequorea* were in 0.5 M NaCl, 5 mM mercaptoethanol, 5 mM EDTA, 0.1% gelatin (w/v), 10 mM Tris–HCl, pH 7.4, 2.5  $\mu$ M coelenterazine (gift from F. McCapra) at room temperature in darkness. Aequorin was reconstituted in whole *Nicotiana* seedlings by floating 8-day-old seedlings in water containing coelenterazine at 1  $\mu$ M in darkness. Before luminescence measurements, seedlings containing reconstituted aequorin were homogenized in 200 mM Tris, 0.5 mM EDTA, pH 7.0, and *in vitro* assays were diluted 1:50 in the same buffer. Reconstituted aequorin was then discharged by the addition of an equal volume of 50 mM CaCl<sub>2</sub> and the total amount of luminescence produced over 10 s measured. Reconstitution in whole *Nicotiana* seedlings was dependent on coelenterazine concentration up to 10  $\mu$ M. Reconstituted aequorin in tobacco was relatively stable, with more than 80% remaining after 48 h. To test the ability of reconstituted aequorin to report  $[Ca^{2+}]_i$ , seedlings were treated with mild detergent and extracellular calcium ( $[Ca^{2+}]_o$ ) to different concentrations. This treatment produced increases in luminescence with increasing  $[Ca^{2+}]_o$  (data not shown).

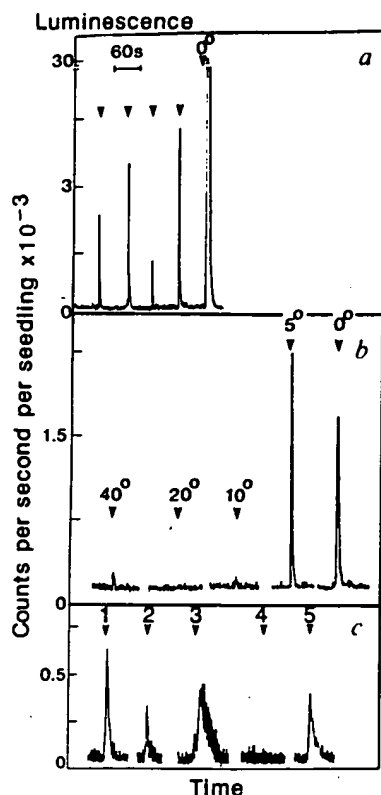


FIG. 3. The effect of touch-simulus (a) temperature-shock (b) and fungal elicitors (c) on calcium-dependent aequorin luminescence.

**METHODS.** Transgenic tobacco seedlings (8 days old) were grown on half-strength M and S medium<sup>25</sup>, 0.8% (w/v) agar. These were then incubated in water containing coelenterazine at 2  $\mu$ M (a, b) or 1  $\mu$ M (c) for 6 h in darkness. For luminescence measurements these seedlings were transferred singly to plastic cuvettes and placed in a luminometer. a, Seedlings were touched once every minute (at arrows) for 4 successive minutes with a fine wire, after which water at 0 °C was added. b, Seedlings were rapidly transferred from 20 °C to temperatures in the range 0–50 °C by the addition of water at the appropriate temperature. Only the treatments at 0 °C, 5 °C, 10 °C, 20 °C and 40 °C are shown. c, The following elicitor preparations were added: (1) untreated yeast elicitor<sup>18</sup>; (2) trifluoroacetic acid (TFA)-hydrolysed yeast elicitor; (3) void volume from a Sephadex G-25 column of yeast elicitor preparation; (4) proteinase K-digested yeast elicitor; and (5) untreated *Glucocladium deliquescens* elicitor<sup>17</sup> (*G. deliquescens* spores a gift from E. Groskurk). TFA hydrolysis was carried out at 120 °C for 1 h with TFA at 2 M. Proteinase K digestion was at 37 °C for 1 h at an enzyme concentration of 1 mg ml<sup>-1</sup>. Luminescence was recorded by a digital luminometer and the output sent to a chart-recorder. All plants tested responded to the stimuli described. Increases in [Ca<sup>2+</sup>]<sub>i</sub> showed low variability between plants in response to reproducible stimuli, for example, cold shock (b), but higher variability in response to unquantifiable stimuli such as touch (a). After measurements seedlings were removed from the luminometer and transferred back to agar, where both vigorous growth and tropic sensitivity were observed. Two other independent tobacco transformants (MAQ2.3 and MAQ2.13; Fig. 1b) showed the same responses (data not shown). No increase in luminescence was detected from untransformed seedlings treated with coelenterazine, nor from untreated transformed seedlings in response to the above stimuli (data not shown).

increase [Ca<sup>2+</sup>]<sub>i</sub> (Fig. 3c). The active fraction in yeast is found in the void volume of a Sephadex-G25 column (Fig. 3c) but not G50 (data not shown); it shows some lability to trifluoroacetic acid under conditions that hydrolyse glycosidic linkages and is completely destroyed by proteinase K (Fig. 3c). Therefore, it is most probably a peptide of relative molecular mass ( $M_r$ ) 5,000–30,000 (5–30K) with attached carbohydrate. Very few fungal elicitors have been characterized although some are carbohydrate and some protein<sup>14</sup>. Characterization of elicitors has so far required assays of enzymes induced in plant cells often over time courses of several days. This method of [Ca<sup>2+</sup>]<sub>i</sub> measurement may greatly simplify elicitor purification and characterization owing to the rapidity of the assay. It has been speculated that elicitor signals are transduced through [Ca<sup>2+</sup>]<sub>i</sub> (ref. 18). Again our data directly support this hypothesis.

The luminescence observed in these experiments (Fig. 3) can only be a qualitative measure of [Ca<sup>2+</sup>]<sub>i</sub> variation. The numbers and types of cells responding cannot be determined and there will be physical barriers (for example, cell walls) to detection of luminescent light from cell layers deeper than two or three cells from the surface. Most probably epidermal cells are the

main contributor. The amount of aequorin discharged in any of these treatments is, however, below 5% of the total determined by homogenization and discharge. It cannot at present be determined whether this discharged aequorin can be regenerated. But the continuous *de novo* synthesis of apoaequorin should allow reconstitution of aequorin several times in the same cell.

We are now analysing the source of the [Ca<sup>2+</sup>]<sub>i</sub> increases described in Fig. 3 by preparing constructs that will target aequorin to plant organelles. The kinetics of the [Ca<sup>2+</sup>]<sub>i</sub> transients induced by touch, cold shock and elicitors are all different. They are very short for touch, intermediate for cold-shock and long for elicitors. Different cellular sources for [Ca<sup>2+</sup>]<sub>i</sub> may explain these variations. Without this considerable simplifying advance in technology these measurements would be very difficult to make. The availability of a range of coelenterazines with different calcium sensitivities<sup>6</sup> can be expected to greatly improve measurements of different organelle calcium and provide additional flexibility to this technique. The goal of monitoring crop plant [Ca<sup>2+</sup>]<sub>i</sub> in the field, which could be used diagnostically, is now brought much nearer with many consequent agricultural benefits. □

Received 24 May; accepted 2 July 1991.

1. Gilroy, S., Hughes, W. A. & Trewavas, A. J. *FEBS Lett.* **199**, 217–221 (1986).
2. Gilroy, S., Read, N. D. & Trewavas, A. J. *Nature* **346**, 769–771 (1990).
3. Shimomura, O., Johnson, F. H. & Saiga, Y. *J. Cell. Physiol.* **59**, 223–240 (1962).
4. Campbell, A. K. *Intracellular Calcium—its Universal Role as Regulator* (Wiley, Chichester, 1983).
5. Williamson, R. E. & Ashley, C. C. *Nature* **296**, 647–651 (1982).
6. Shimomura, O., Musicki, B. & Kishi, Y. *Biochem. J.* **261**, 913–920 (1989).
7. Prasher, D. C., McCann, R. O. & Cormier, M. J. *Biochem. biophys. Res. Commun.* **126**, 1259–1268 (1985).
8. Franck, G., Guille, M., Jonard, G., Richards, K. & Hirth, L. *Cell* **21**, 285–294 (1980).
9. Bevan, M. *Nucleic Acids Res.* **12**, 8711–8721 (1984).
10. Blinks, J. R. et al. *Meth. Enzym.* **57**, 292–328 (1978).
11. Braam, J. & Davis, R. W. *Cell* **60**, 357–364 (1990).
12. Minorsky, P. V. *Pl. Cell Envir.* **8**, 75–94 (1985).
13. Mauch, F. & Staehelin, L. A. *Pl. Cell* **1**, 447–457 (1989).
14. Dixon, R. A. *Biol. Rev.* **63**, 239–291 (1988).
15. Watson, D. G., Rycroft, D. S., Freer, I. M. & Brooks, C. J. M. *Phytochemistry* **24**, 2195–2200 (1985).

16. Hahn, M. G. & Albersheim, P. *Pl. Physiol.* **62**, 107–111 (1978).
17. Brooks, C. J. W., Watson, D. G. & Freer, I. M. *Phytochemistry* **25**, 1089–1092 (1986).
18. Ebel, J. & Grisebach, H. *Trends biochem. Sci.* **13**, 23–27 (1988).
19. Prasher, D. C., McCann, R. O., Longiaru, M. & Cormier, M. in *Bioluminescence and Chemiluminescence—New Perspectives* (ed. Scholmerich, J.) 365–368 (Wiley, Chichester, 1986).
20. Pietrzak, M., Shillito, R. D., Hohn, T. & Potrykus, I. *Nucleic Acids Res.* **14**, 5857–5868.
21. Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, New York, 1989).
22. Draper, J., Scott, R., Armitage, P. & Walden, R. *Plant Genetic Transformation and Gene Expression—A Laboratory Manual* (Blackwell, Oxford, 1988).
23. Harlow, E. & Lane, D. *Antibodies—A Laboratory Manual* (Cold Spring Harbor Laboratory Press, New York, 1988).
24. Campbell, A. K., Patel, A. K., Razavi, Z. S. & McCapra, F. *Biochem. J.* **252**, 143–149 (1988).
25. Murashige, T. & Skoog, F. *Pl. Physiol.* **15**, 473–497 (1962).

**ACKNOWLEDGEMENTS.** We acknowledge financial support by the SERC and photographic assistance from W. Sinclair.

# Expression of a Maize Cell Wall Hydroxyproline-Rich Glycoprotein Gene in Early Leaf and Root Vascular Differentiation

Virginia Stiefel,<sup>a,b</sup> Luis Ruiz-Avila,<sup>a</sup> Regina Raz,<sup>a</sup> Maria Pilar Vallés,<sup>a</sup> Jordi Gómez,<sup>a</sup> Montserrat Pagés,<sup>a</sup> Jose Antonio Martínez-Izquierdo,<sup>a</sup> Maria Dolores Ludevid,<sup>a</sup> Jane A. Langdale,<sup>b</sup> Timothy Nelson,<sup>b</sup> and Pere Puigdomènech<sup>a,1</sup>

<sup>a</sup> Departamento de Genética Molecular, CID-CSIC, Jordi Girona 18, 08034 Barcelona, Spain

<sup>b</sup> Department of Biology, Yale University, New Haven, Connecticut 06511

The spatial pattern of expression for a maize gene encoding a hydroxyproline-rich glycoprotein (HRGP) was determined by *in situ* hybridization. During normal development of roots and leaves, the expression of the gene was transient and particularly high in regions initiating vascular elements and associated sclerenchyma. Its expression was also associated with the differentiation of vascular elements in a variety of other tissues. The gene encoded an HRGP that had been extracted from the cell walls of maize suspension culture cells and several other embryonic and post-embryonic tissues. The gene was present in one or two copies in different varieties of maize and in the related monocots teosinte and sorghum. A single gene was cloned from maize using a previously characterized HRGP cDNA clone [Stiefel et al. (1988). *Plant Mol. Biol.* 11, 483-493]. In addition to the coding sequences for the HRGP and an N-terminal signal sequence, the gene contained a single intron in the nontranslated 3' end.

## INTRODUCTION

The specialization of cell wall architecture is an important feature of the functional differentiation of plant cells (Varner and Lin, 1989). For example, the rigidity of sclerenchyma cells, the pressure-resistant nature of xylem cells, and the gas impermeability of photosynthetic bundle sheath cells all rely on the properties of specialized walls that each of these cells deposits during differentiation. The wall of each distinct cell type appears to have a characteristic combination and spatial organization of polysaccharides, structural proteins, and other wall components, often with unique patterns of cross-links and other modifications (Fry, 1986; Cassab and Varner, 1988). These structural specializations are accomplished in part through the differential expression of genes encoding wall structural proteins and synthetic and modification enzymes.

The cell-specific expression of several genes encoding specialized wall components has been described in dicots. A gene encoding phenylalanine ammonia-lyase, the enzyme catalyzing the first step in the synthesis of lignin monomers, is expressed at developing vascular centers, coincident with the differentiation of lignified xylem elements (Bevan et al., 1989; Liang et al., 1989). Similarly, a gene encoding a member of the wall glycine-rich protein

(GRP) class that is accumulated in lignified secondary walls of xylem elements is expressed specifically in differentiating protoxylem cells (Keller et al., 1989a). The genes for several members of another class of wall-associated proteins, the hydroxyproline-rich glycoproteins (HRGPs), have been shown to have distinct cellular patterns of expression.

The most studied of the HRGPs are the dicot extensins (Cassab and Varner, 1988; Showalter and Varner, 1989). Extensins contain a characteristic repeat of the pentapeptide Ser-Pro<sub>4</sub>, in which proline residues are hydroxylated and glycosylated. Extensins become insoluble with time and contribute to the mechanical strength of the wall, probably via cross-linking to other monomers and to other wall components (Fry, 1986). Extensins have been localized in dicots to the walls of several cells in which strength is a key property, including seed coat sclerenchyma and cotyledon vascular elements (Cassab and Varner, 1987). Individual extensin genes are expressed in tomato and tobacco with temporal and spatial patterns that suggest a high degree of developmental control (Showalter et al., 1985; Memelink et al., 1987). Extensin accumulation can also be induced by fungal infection and by wounding, although distinct extensin genes may be induced in each case (Showalter et al., 1985; Corbin et al., 1987). Recently,

<sup>1</sup> To whom correspondence should be addressed.

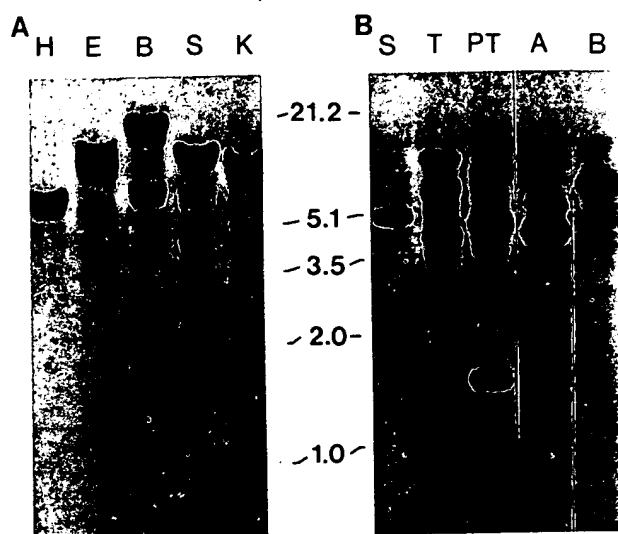
a distinct tobacco HRGP was described that is specifically synthesized and accumulated at sites of lateral root initiation (Keller and Lamb, 1989).

Here we describe the expression pattern of a gene encoding a maize HRGP. We reported previously the isolation of a cDNA encoding a protein consisting of 13 repeats of a proline- and threonine-rich peptide (Stiefel et al., 1988). This protein contains only a single copy of the Ser-Pro<sub>4</sub> motif repeated in extensins. The deduced protein probably corresponds to the hydroxyproline- and threonine-rich glycoproteins extracted from maize pericarp (Hood et al., 1988) and cell walls of suspension culture cells and various seedling and embryo tissues (Kieliszewski and Lamport, 1987; Kieliszewski et al., 1990). The correspondence of these protein preparations to the protein predicted by cDNA sequencing was recently confirmed by peptide sequencing (Kieliszewski et al., 1990). The gene encoding this protein is present in one or two copies in the genomes of maize, teosinte, and sorghum. The maize transcript is interrupted by a single intron in the 3' untranslated region. We show that the gene is expressed at sites of early vascular differentiation in embryos, coleoptiles, leaves, hypocotyl, and both primary and lateral roots, as well as at much lower levels throughout the developing plant.

## RESULTS

### Maize HRGP Is Encoded by a Gene with a 3' Intron

We previously described the isolation of a cDNA encoding a maize HRGP and showed that the corresponding mRNA is enriched in tissues with mitotic activity (Stiefel et al., 1988) and in wounded tissues (Ludevid et al., 1990). The extensins, a group of HRGP proteins that have been found in differentiating and wounded tissues of several dicot species (Showalter and Varner, 1989), are encoded by small gene families. We performed genomic blot analysis to determine the number of sequences in the maize genome and in those of the monocots teosinte and sorghum. Genomic DNA of maize, teosinte, and sorghum was digested with a variety of restriction enzymes, blotted to nylon membrane, and probed with the maize HRGP cDNA (pMC56), all as described in Methods. Figures 1A and 1B show that the HRGP gene is present in one or two copies in the genomes of four different maize inbred lines. Using the same cDNA probe, the gene has been mapped to a single locus on maize chromosome 2 by RFLP analysis (locus UMC 145, C. Guitton and D. Hoisington; personal communication). Consistent with this, independent cDNAs from plumule (leaf) and root RNA were found to have identical sequences in noncoding regions, suggesting that these different tissues contain transcripts from a single gene (V. Stiefel, unpublished results). Figure 1B shows



**Figure 1.** Genomic DNA Blot Analysis of Sequences Homologous to the HRGP cDNA in Maize and Related Species.

(A) DNA (10  $\mu$ g) from the maize W64A variety was digested with HindIII (H), EcoRI (E), BamHI (B), SacI (S), and KpnI (K).

(B) DNA (10  $\mu$ g) from different sources was digested with SacI. Samples are from sorghum (S), teosinte (T), and the maize varieties Palomero Toluqueno (PT), A188 (A), and Black Mexican Sweet (B). In both cases HRGP cDNA (probe MC56 from Stiefel et al., 1988) was used as a probe. Position of size markers (in kb) is shown.

that the teosinte and sorghum genomes contain similar sequences, also in relatively low copy number.

We cloned a genomic copy of the sequences represented in the HRGP cDNA, as described in Methods. Several overlapping clones were isolated from a maize inbred AC1503 library in  $\lambda$ EMBL3, and one was chosen for sequence analysis. The DNA sequence of 1.8 kb including the HRGP gene is shown in Figure 2. A TATA box and polyadenylation signal are found at expected locations in 5' and 3' regions flanking the HRGP coding region at -112 bp and +1492 bp relative to the initial ATG. Upstream of the TATA box, several short repetitive sequences and polypyrimidine-polypurine stretches are found.

Comparison of the HRGP genomic sequence with the previously determined cDNA sequence (Stiefel et al., 1988) revealed some minor sequence differences in the 3' untranslated region, probably due to the difference in maize inbred varieties used for cDNA and genomic cloning (W64A versus AC1503). Comparison of sequences in the coding region indicated that the characterized maize cDNA represents an mRNA from this gene. In addition, the genomic HRGP clone includes a 166-bp sequence (underlined in Figure 2) delimited by consensus sequences for intron splice junctions. An extensin gene from carrot has this

agggcatcggaggccccccccccccc -121  
 cctccgtgtatataagcagtgagggtgagcgtctctcctcagaccactgcgccat -61  
 tggccagctagagccaaccagaagagcttgagcttactgagagtggtgtgagagagagg -1  
 ATGGGTGGCAGCGGCGAGGCTGCTCTGCTGCTGCGCTGGTGGTGGCGGTGAGCGTG 60  
 M G G S G R A A L L L A L V V V A V S L 20  
 CCGGTGGAGATCCAGCGCGACGCGGGTACGGTTACGGCGGGGTACACCGACGCGG 120  
 A V E I Q A D A G Y G G G Y T P T P 40  
 ACGCGGCGACCGCGACCGGAGCGCGGAGCGCGGCGCGGCGCGGCGCGGCGGCGG 180  
 T P A T P T P K P E K P P T K G P K P D 60  
 AAGCGCGGAGGAGCGCGGCGGAGCGCGGAGCGCGGCGCGGCGCGGCGCGGCGG 240  
 K P P K E H G P K P P K E H K P T 80  
 CCGCGGAGTACACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 300  
 P P T Y T P S P K P T P P T Y T P T P 100  
 CCGCGGAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 360  
 P P K P T P T P P T P P T P H K P T P 120  
 AAACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 420  
 K P T P T P P T Y T P T P K P P T P 140  
 ACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 480  
 K P P T Y T P S P K P P T P K P T P P T 160  
 TACACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 540  
 Y T P S P K P P T P K P T P P T Y T P S 180  
 CCTAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 600  
 P K P P A T P P T P P T P P T P P T Y T P 200  
 TCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 660  
 S P K P P T P K P T P P T Y T P S P K P 220  
 CCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 720  
 P T K P T P T P P T P P T P P T P P T H P 240  
 ACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 780  
 T P K P T P P T Y T P S P K P P T P K P 260  
 ACCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 840  
 T P P T Y T P S P K P P T P K P T P P T 280  
 TACACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 900  
 Y T P S P K P P T P K P T P P T Y T P T 300  
 CGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960  
 P K P P A T P P T Y T P T P P V S H T 320  
 CCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1020  
 P S P P P P Y Y 328  
 tctctggagcatttaggtacgtactactacgtatcgtacagaatggagcatgcaa 1080  
 tggatgacacactgacatatttagtatgtgtgtgcaaatgtatcgtcagtatcat 1140  
 actgatctcctggcatagctcggcactaacataggctctctcttcttctgtgtggac 1200  
 aggtggtcctggatcaatggaagggtgtgtcctagccagcggcgaagatgagctgctga 1260  
 tggtaatgatgatgaagagaccactgctacgtacctctctcttctgtgtgtgctc 1320  
 cgtcccgctagacgatcagagagagaatagcagagctctgtgtccggcctctgtctt 1380  
 ctcgctccggcgcttaatttactagtgtgtgtccttatatgttttagcagcagcagg 1440  
 tgtattgtcgggatgtatgtattgcaactatattgggtgtaaaaccatataaattg 1500  
 tgggcaaatatgcaagaaaaaggcttctactctactgtacacttactacc 1553

**Figure 2.** Nucleotide Sequence of the Maize HRGP Gene and the Corresponding Protein Sequence.

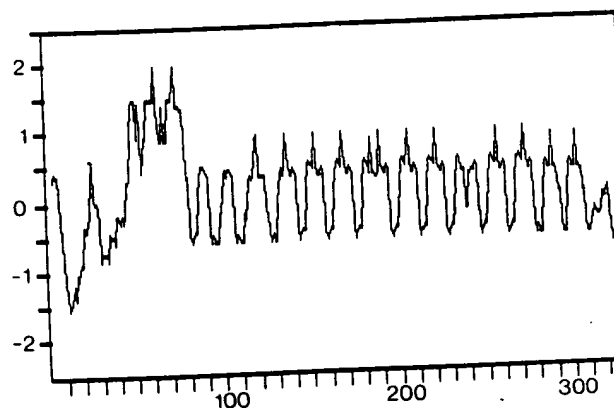
Nucleotides are numbered from the A of the start codon for translation (+1). Putative TATA box and polyadenylation signals are boxed. The intron in the 3' untranslated end is underlined. The deduced amino acid sequence for HRGP is depicted as single letter code. The site of processing of the mature protein between Ala 26 and Asp 27 is indicated by a vertical arrow.

same unusual feature of an intron in the 3' untranslated region of the transcript (Chen and Varner, 1985).

The protein predicted by the genomic sequence was identical to that predicted by the cDNA, with the additional

N-terminal amino acids that were missing in the (incomplete) cDNA. Figure 3 shows that the predicted protein begins with a hydrophobic stretch having features typical of a signal peptide. As confirmed by N-terminal amino acid sequencing (see below), the mature protein sequence begins with the aspartic acid residue in position 27, followed by a short stretch (8 residues) of glycine and tyrosine residues. This is followed by a proline-rich region that includes the repeated peptide Gly-Pro-Lys-Pro-(Asp/Glu)-Lys-Pro-Pro-Lys-Glu-His. Finally, the proline-, threonine-rich region that corresponds to the sequenced cDNA begins. The repeated hydrophilic/hydrophobic character of the unmodified mature protein is very apparent in this hydropathy plot.

We identified the N-terminal amino acid of the mature protein by N-terminal amino acid sequencing of the protein encoded by the maize HRGP gene. A cell wall protein with characteristics of the protein predicted by the above sequence has been isolated from maize pericarp (Hood et al., 1988), suspension culture cells (Kieliszewski and Lamport, 1987), and several other embryo and seedling tissues (Kieliszewski et al., 1990). The amino acid composition of these proteins matches that predicted from the HRGP DNA sequence (Table 1), and Kieliszewski et al. (1990) confirm this with partial amino acid sequencing. We used a similar purification scheme (see Methods) to isolate the corresponding protein fraction from maize coleoptiles. This protein has the amino acid composition (Table 1) and repeated chymotryptic cleavage pattern (data not shown) expected for the encoded HRGP. As reported by Kieliszewski et al. (1990), the extracted protein migrates as a diffuse band at approximately 70 kD on SDS polyacrylamide gels, presumably due to its extensive modification and high proline content. We determined the N-terminal



**Figure 3.** Hydropathy Profile of the Amino Acid Sequence Deduced from the Genomic Clone of Maize HRGP.

The mean hydropathy of a window of 6 consecutive residues (as described by Hopp and Wood, 1981) is plotted against the amino acid number.

**Table 1.** Amino Acid Composition (in mol %) of Maize HRGPs from Different Sources

	E <sup>a</sup>	P <sup>b</sup>	PC-1 <sup>c</sup>	THRGP <sup>d</sup>
HP <sup>e</sup>		24.7	21.9	24.8
P	45.4	11.7	13.5	14.5
T	23.8	24.0	17.5	25.3
K	12.2	15.0	11.3	13.5
Y	6.6	4.6	4.6	3.9
S	4.0	4.7	5.5	7.3
G	2.3	5.1	7.1	2.4
A	1.7	5.7	5.2	1.7
E	1.3	2.3	2.5	2.3
H	1.7	1.8	3.6	2.4
D	0.7	2.4	—	0.7
V	0.3	0.9	2.7	0.7
Others	—	1.0	4.6	0.5

<sup>a</sup> Expected amino acid composition deduced from the nucleotide sequence of the HRGP gene. Amino acids corresponding to signal peptide (residues 1 to 26) were not computed.

<sup>b</sup> HRGP protein purified from maize coleoptiles as described in Methods.

<sup>c</sup> HRGP protein purified from maize pericarps (Hood et al, 1988).

<sup>d</sup> HRGP protein from Black Mexican Sweet suspension cultures (Kieliszewski and Lampert, 1987).

<sup>e</sup> HP, hydroxyproline.

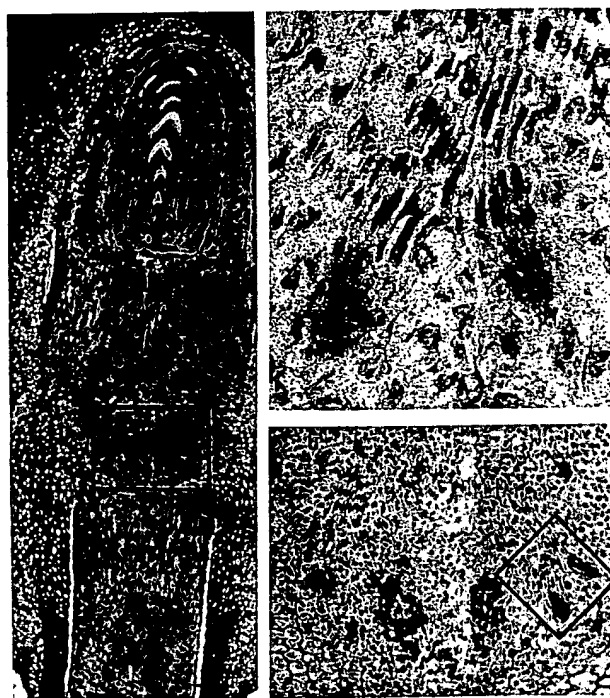
sequence of the extracted protein by Edman degradation with a gas phase sequencer. Sequencing cycles 4, 6, and 10 gave tyrosine residues followed by the sequence Thr-Pro-Thr-Pro-X-Pro-Ala. This sequence appears at amino acids 37 to 43 of the genomic sequence open reading frame. Residues at positions 1 to 3, 5, and 7 to 9 (corresponding to glycine) relative to the N terminus were not clearly assigned due to a high background. From the amino acid sequence predicted by the genomic clone, this indicates that amino acid 1 is aspartate. The predicted hydrophobic sequence prior to this aspartate corresponds to a signal peptide that is probably processed at the Ala-Asp junction between positions 26 and 27 of the predicted protein. This would make the mol wt of the processed protein (without modifications) 31,729 D.

#### HRGP Transcripts Are Localized at Sites of Vascular Differentiation

Our previous work suggests that mRNA for the maize HRGP and extractable HRGP are enriched in a variety of tissues coincident with mitotic activity during normal development or after wounding (Stiefel et al., 1988; Ludevid et al., 1990). HRGP mRNA abundance decreases dramatically in mature tissue. To localize more precisely the accumulation of HRGP mRNA, we used cDNA pMC56 as a probe for in situ hybridizations with developing and mature tissues of the shoot and root systems. Figures 4 to 7 show that mRNA was localized to regions of vascular

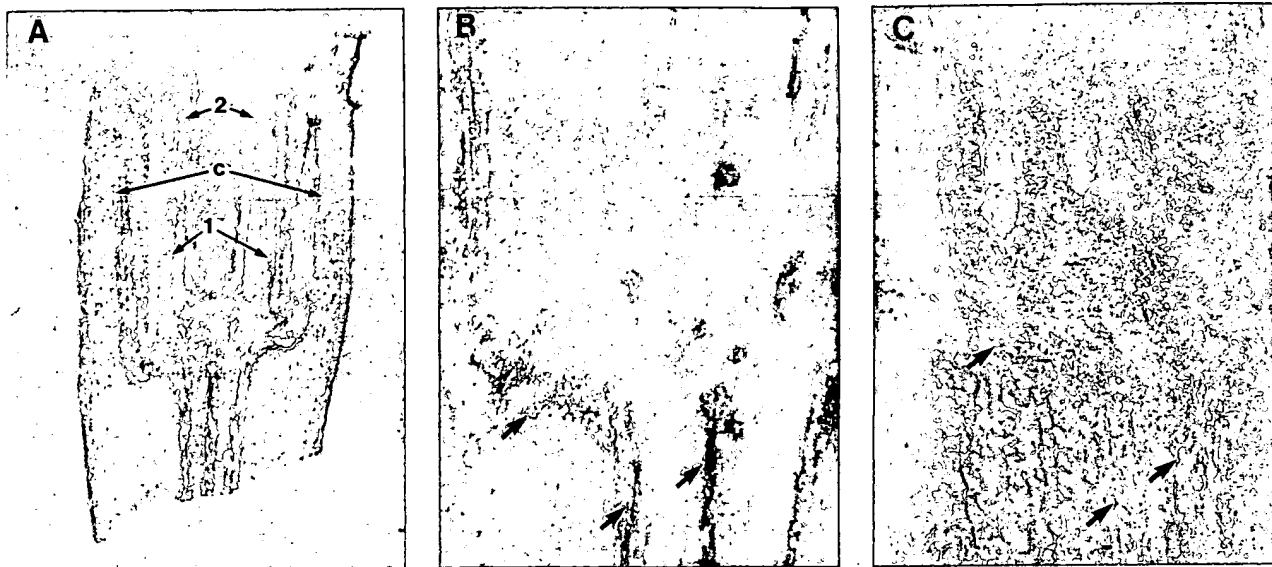
differentiation in hypocotyl and mesocotyl regions of the embryo axis (1 day post-germination), the first node of the developing shoot, leaf primordia in the plumule (6 days post-germination), and sheaths of more mature developing leaves. The resolution of the hybridization technique does not permit the assignment of signals over specific vascular elements, although signals are strongest over xylem elements and surrounding sclerenchyma. In each case, the most abundant signal appeared at a morphological stage in which xylem tracheary element differentiation is in progress.

HRGP mRNA accumulated transiently in regions undergoing vascularization in developing primary and lateral roots. Figure 8 shows an oblique section through the tip region of a primary root. Hybridization was strongest in regions of differentiating metaxylem and protoxylem elements. The region of hybridization was always localized in a differentiating zone near the tip and was absent in more distant (more mature) regions. Figure 9 shows the highly localized accumulation of mRNA at sites of lateral root initiation, just behind the advancing lateral tip. Hybridization appeared to be most intense in regions of vascular differentiation and formed a continuous connection from the primary root vascular cylinder to the new lateral root



**Figure 4.** Localization of HRGP Transcripts in the Maize Embryo Axis.

Sections of 1 day post-germination embryos were hybridized with the HRGP probe as described in Methods.



**Figure 5.** Localization of HRGP Transcripts in the Region of the Coleoptilar Node.

Longitudinal sections of 7 day post-germination plants were hybridized with the HRGP probe.

(A) Coleoptilar node region. c, coleoptile; 1, first leaf; 2, second leaf.

(B) Enlargement of same section, hybridized with antisense strand (signal) probe. Arrows indicate signal over coleoptilar vein and central vein in axis.

(C) Adjacent section hybridized with sense strand (control) probe. Arrows indicate absence of signal at same locations.

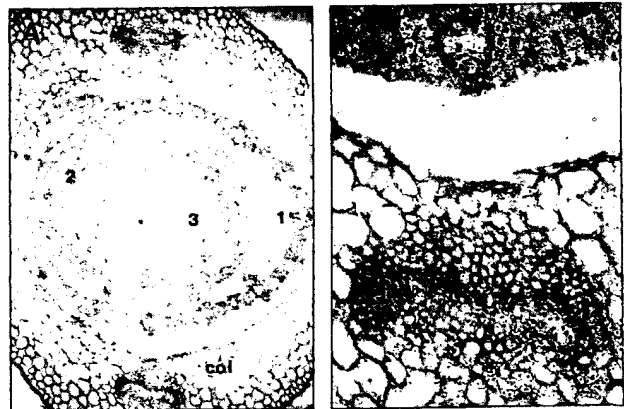
tip. This cross-section also shows that no significant hybridization occurred over the now mature meta- and protoxylem elements of the primary root.

acidic and basic amino acids, a feature also observed in putative cell wall proteins from dicots (Franssen et al., 1987; Hong et al., 1987). The remaining 248 amino acids

## DISCUSSION

We have determined the DNA sequence and expression pattern for a maize gene encoding a highly repetitive proline- and threonine-rich protein. The deduced amino acid sequence suggests that it corresponds to an HRGP that can be extracted from cell walls of several maize tissues (Kieliszewski and Lamport, 1987; Hood et al., 1988; Kieliszewski et al., 1990).

The predicted protein primary structure included several distinct blocks that may have distinct wall structural functions. A hydrophobic signal sequence was located at the N terminus and was absent in the mature protein, presumably due to processing that accompanies its passage through the endoplasmic reticulum to the wall (Gardiner and Chrispeels, 1975; Von Heijne, 1981). The mature protein began with an eight amino acid hydrophobic glycine-tyrosine-rich region. At least one dicot glycine-rich protein is highly specific for tracheary element walls (Keller et al., 1989b). The glycine-rich region was followed by a short stretch of proline-rich sequence with alternating

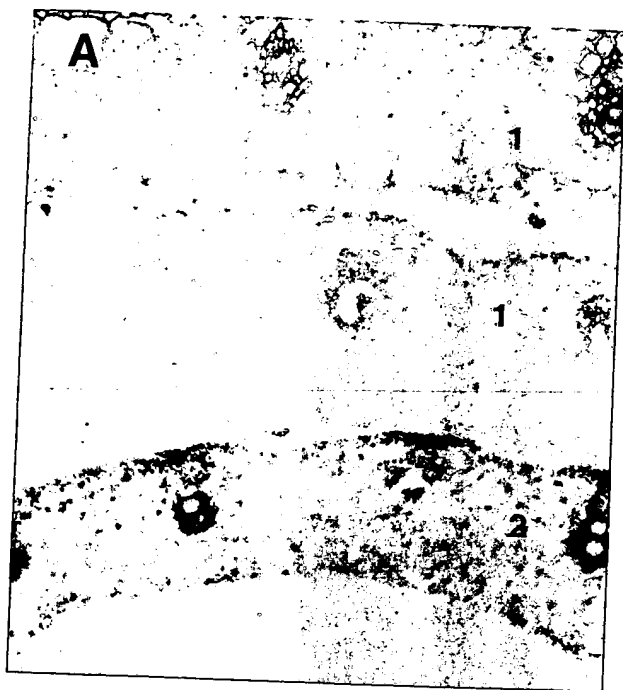


**Figure 6.** Localization of HRGP Transcripts in Plumules.

Cross-sections through plumules of 6 day post-germination plants were hybridized with the HRGP probe.

(A) Plumule cross-section. col, coleoptile; 1, 2, 3, leaves 1, 2, and 3. Signals are present at coleoptilar veins and at major veins of leaf 1.

(B) Enlargement of region including coleoptilar vein and adjacent leaf 2 vein. Signals over both veins indicated by arrows.



**Figure 7.** Localization of HRGP Transcripts in Sheaths of Developing Leaves.

Cross-sections through 14 day post-germination shoots at the level of the developing sheaths of leaves 1 and 2 were hybridized with the HRGP probe.

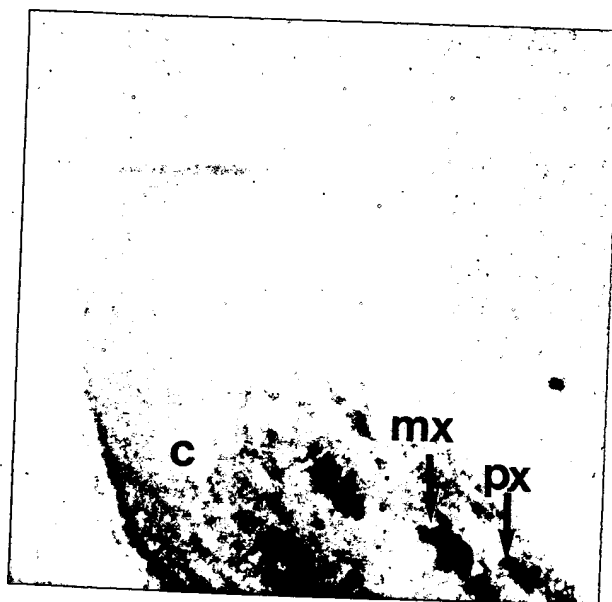
(A) Sheath regions of leaves 1 (1) and 2 (2). Signals are present only over veins of leaf 2.

(B) Enlargement of vein from leaf 2. Signal is apparent over xylem and adjacent sclerenchyma (arrows).

formed blocks of repetitive sequences in which tyrosine residues were regularly located in the most hydrophobic part of the molecule (Stiefel et al., 1988). Tyrosine and threonine residues were predominantly located in PPTY peptides, a motif that is repeated in a tobacco HRGP associated with lateral root initiation (Keller and Lamb, 1989). Tyrosine residues have been proposed as sites of extensin polymerization (Fry, 1986). The hydrophobic environment of regularly spaced tyrosines in this maize HRGP might permit a polymerized hydrophobic surface on one face with a hydrophilic opposite surface, although this pattern would be altered by modifications such as glycosylations, proline hydroxylations, and tyrosine polymerization. Near the C terminus is a single Ser-Pro<sub>4</sub> sequence which is a motif repeated in extensins (Showalter and Varner, 1989).

The localization of the maize HRGP mRNA and the corresponding protein suggests that it plays a role in the early construction of walls surrounding a vascular element or elements common to many organs. We showed that mRNA accumulates transiently at new vascular sites in embryos, leaves, and roots. This pattern is similar to that observed in bean for a glycine-rich protein (GRP), which appears to be a component of xylem walls (Keller et al., 1988). This maize HRGP mRNA is specifically induced at wound sites in young tissues (Ludevid et al., 1990). This pattern may represent the regeneration of vascular elements to circumvent wounded sites, although the rapid time course (mRNA peak at 1 hr to 2 hr) makes it less likely.

The highly localized accumulation of the mRNA at sites



**Figure 8.** Localization of HRGP Transcripts in the Primary Root.

Oblique sections through tip region of 10 day post-germination root systems were hybridized with the HRGP probe. Arrows indicate hybridization signal over differentiating meta- (mx) and protoxylem (px) elements. c, cortex.





**Figure 9.** Localization of HRGP Transcripts in Developing Lateral Roots.

Cross-sections through 10 day post-germination root systems were hybridized with the HRGP probe. This section through the primary root is at the level of a newly initiated lateral root. c, cortex; e, endodermis; pc, pericycle; px, protoxylem; mx, metaxylem. No signal is present in primary root. Arrow indicates considerable signal over developing vascular bundle of lateral root.

of lateral root initiation might also be interpreted as being at sites of wounding, as the initiated lateral root penetrates existing tissue. However, our in situ hybridization patterns showed that the HRGP mRNA is associated with the developing vascular link of the new lateral root to the primary root, rather than at the "wounding" tip. This pattern is in contrast to the accumulation pattern of a distinct HRGP gene in tobacco, which is expressed only at the tips of new lateral roots and which may have a mechanical

role in penetration of new roots through existing tissue (Keller and Lamb, 1989). However, we also observed accumulation of the maize HRGP in regions associated with xylem differentiation in the primary root, where no wound-like process occurs. The distinct patterns of gene expression and protein accumulation for various dicot and monocot HRGPs suggest that they constitute a family of proteins with diverse roles in plant walls.

## METHODS

### Genomic Blot Analysis

Genomic DNA was prepared from leaves of maize inbreds W64A, A188, Palomero Toluqueño, and Black Mexican Sweet varieties and from teosinte (*Zea diploperennis*) and sorghum (*Sorghum bicolor*), as described by Burr and Burr (1981). The DNA was digested with restriction enzymes, separated by size in 0.8% agarose gels, and blotted onto nylon membranes as recommended by the manufacturer (Zeta-Probe, Bio-Rad). Hybridization was carried out in 180 mM NaCl, 10 mM sodium phosphate, 10 mM EDTA, 1% SDS, 0.5% nonfat milk, 0.5 mg/mL sonicated salmon sperm DNA at 68°C. The DNA probe was labeled with  $^{32}\text{P}$  to a specific activity of 0.5 to  $2 \times 10^9$  cpm/ $\mu\text{g}$  by random priming (Boehringer Mannheim). Final washes were done in  $0.1 \times \text{SSC}$ , 0.1% SDS at 65°C, and the membrane was exposed to Kodak XAR5 film with intensifying screens (DuPont Lightning Plus) at -70°C.

### Genomic Cloning

A maize inbred AC1503 genomic library (kindly provided by A. Gierl, Köln; Sau3A partial library in  $\lambda\text{EMBL3}$ ) was screened with the insert of the pMC56 HRGP cDNA clone (Stiefel et al., 1988) by standard methods (Sambrook et al., 1989). A 5-kb genomic clone hybridizing with the probe was sequenced using the dideoxy method after subcloning in M13mp18 and 19 (Sanger et al., 1977); 100% of both strands were sequenced over the 5-kb region.

### Protein Analysis

Protein for amino acid analysis and N-terminal sequencing was extracted (Mazeau et al., 1982) from the coleoptiles of 6-day-old maize plantlets (inbred W64A) and deglycosylated as previously described (Stiefel et al., 1988). Amino acid analyses of purified HRGP were carried out after acid hydrolysis in a Pico-Taq (Millipore) analyzer according to the manufacturer. N-terminal amino acid sequence was determined in a gas-phase sequencer (model 470A, Applied Biosystems) according to the manufacturer.

### In Situ Hybridization

For post-embryonic material, histological techniques and in situ hybridization were performed on paraffin-embedded samples as previously described (Langdale et al., 1988). Frozen sections were

used for embryos. Embryos germinated for 48 hr were embedded and frozen in OTC (Tissue Tek) compound on dry ice. Cryostat (Reichert-Jung) sections 8  $\mu$ m to 10  $\mu$ m thick were collected on gelatin-subbed slides, dried on a hot plate at 60°C for 1 min, and fixed in a freshly made 4% paraformaldehyde solution in PBS (0.15 M NaCl, 10 mM sodium phosphate, pH 7.4). <sup>35</sup>S-labeled riboprobes were synthesized with T3 and T7 RNA polymerases (Bethesda Research Laboratories), using linearized cDNA subclones derived from pMC56 as template.

## ACKNOWLEDGMENTS

The authors are indebted to Dr. Alphons Gierl for the gift of his maize genomic library, to M. Patricio Gómez for assistance in DNA sequencing, and to Drs. Jean-Jacques Leguay and Jean-Claude Guillemot (Sanofi Elf-BioRecherches, Toulouse) for determination of N-terminal sequence. L. R.-A., R.R., M.P.V., and J.G. are recipients of Plan de Formación Investigador fellowships, V.S. is the recipient of a Fulbright Fellowship, and L.R.-A. is the recipient of a short-term Comissió Interdepartamental de Recerca i Innovació Tecnològica fellowship. The work has been supported by grants from Plan Nacional de Investigación Científica y Técnica (BIO 88/0242) and from the European Communities (BAP-374) to P.P.

Received May 25, 1990; revised June 8, 1990.

## REFERENCES

- Bevan, M., Shufflebottom, D., Edwards, K., Jefferson, R., and Schuch, W. (1989). Cell- and tissue-specific activity of a phenylalanine ammonia-lyase promoter in transgenic plants. *EMBO J.* **8**, 1899–1906.
- Burr, B., and Burr, F.A. (1981). Controlling-element events at the shrunken locus in maize. *Genetics* **98**, 143–156.
- Cassab, G.I., and Varner, J.E. (1987). Immunocytochemical localization of extensin in developing soybean seed coats by immunogold-silver staining and by tissue printing on nitrocellulose paper. *J. Cell Biol.* **105**, 2581–2588.
- Cassab, G.I., and Varner, J.E. (1988). Cell wall proteins. *Annu. Rev. Plant Physiol.* **39**, 321–353.
- Chen, J., and Varner, J.E. (1985). An extracellular matrix protein in plants: Characterization of a genomic clone for carrot extensin. *EMBO J.* **4**, 2145–2151.
- Corbin, D.R., Sauer, N., and Lamb, C.J. (1987). Differential regulation of a hydroxyproline-rich glycoprotein gene family in wounded and infected plants. *Mol. Cell. Biol.* **7**, 4337–4344.
- Franssen, H.J., Nap, J.P., Gloudemans, T., Stiekema, W., van Dam, H., Govers, F., Louwerse, J., van Kammen, A., and Bisseling, T. (1987). Characterization of a cDNA for nodulin-75 of soybean: A gene product involved in early stages of root nodule development. *Proc. Natl. Acad. Sci. USA* **84**, 4495–4499.
- Fry, S.C. (1986). Cross-linking of matrix polymers in the growing cell walls of angiosperms. *Annu. Rev. Plant Physiol.* **37**, 165–186.
- Gardiner, M., and Chrispeels, M.J. (1975). Involvement of the Golgi apparatus in the synthesis and secretion of hydroxyproline-rich cell wall glycoproteins. *Plant Physiol.* **55**, 536–541.
- Hong, J.C., Nagao, R.T., and Key, J.L. (1987). Characterization and sequence analysis of a developmentally regulated putative cell wall protein gene isolated from soybean. *J. Biol. Chem.* **262**, 8367–8376.
- Hood, E.E., Shen, Q.X., and Varner, J.E. (1988). A developmentally regulated hydroxyproline-rich glycoprotein in maize pericarp cell walls. *Plant Physiol.* **87**, 138–142.
- Hopp, T.P., and Wood, K.R. (1981). Prediction of protein antigenic determinants from amino acid sequences. *Proc. Natl. Acad. Sci. USA* **78**, 3824–3828.
- Keller, B., and Lamb, C.J. (1989). Specific expression of a novel cell wall hydroxyproline-rich glycoprotein gene in lateral root initiation. *Genes Dev.* **3**, 1639–1646.
- Keller, B., Sauer, N., and Lamb, C.J. (1988). Glycine-rich cell wall proteins in bean: Gene structure and association of the protein with the vascular system. *EMBO J.* **7**, 3625–3633.
- Keller, B., Schmid, J., and Lamb, C.J. (1989a). Vascular expression of a bean cell wall glycine-rich protein- $\beta$ -glucuronidase gene fusion in transgenic tobacco. *EMBO J.* **8**, 1309–1314.
- Keller, B., Templeton, M.D., and Lamb, C.J. (1989b). Specific localization of a plant cell wall glycine-rich protein in protoxylem cells of the vascular system. *Proc. Natl. Acad. Sci. USA* **86**, 1529–1533.
- Kieliszewski, M., and Lamport, D.T.A. (1987). Purification and partial characterization of a hydroxyproline-rich glycoprotein in a graminaceous monocot, *Zea mays*. *Plant Physiol.* **85**, 823–827.
- Kieliszewski, M., Leykam, J.F., and Lamport, D.T.A. (1990). Structure of the threonine-rich extensin from *Zea mays*. *Plant Physiol.* **92**, 316–326.
- Langdale, J.A., Rothermel, B., and Nelson, T. (1988). Cellular pattern of photosynthetic gene expression in developing maize leaves. *Genes Dev.* **2**, 106–115.
- Liang, X., Dron, M., Schmid, J., Dixon, R.A., and Lamb, C.J. (1989). Developmental and environmental regulation of a phenylalanine ammonia-lyase- $\beta$ -glucuronidase gene fusion in transgenic tobacco plants. *Proc. Natl. Acad. Sci. USA* **86**, 9284–9288.
- Ludevid, M.D., Ruiz-Avila, L., Vallés, M.P., Stiefel, V., Torrent, M., Torné, J.M., and Puigdomènech, P. (1990). Expression of cell wall protein genes in dividing and wounded tissues of *Zea mays*. *Planta* **180**, 524–529.
- Mazeau, D., Rumeau, D., and Esquerré-Tugayé, M.T. (1982). Two different families of hydroxyproline-rich glycoproteins in melloid callus: Biochemical and immunochemical studies. *Plant Physiol.* **86**, 540–546.
- Memelink, J., Hoge, J.H.C., and Schilperoort, R.A. (1987). Cytokinin stress changes the developmental regulation of several defence-related genes in tobacco. *EMBO J.* **6**, 3579–3583.
- Sambrook, J., Fritsch, E.F., and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*. Second Edition. (Cold Spring

- Harbor, NY: Cold Spring Harbor Press).
- Sanger, F., Nicklen, S., and Coulson, A.R.** (1977). DNA sequencing with chain-termination inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.
- Showalter, A.M., Bell, J.N., Cramer, C.L., Bailey, J.A., Varner, J.E., and Lamb, C.J.** (1985). Accumulation of hydroxyproline-rich glycoprotein mRNAs in response to fungal elicitor and infection. *Proc. Natl. Acad. Sci. USA* **82**, 6551-6555.
- Showalter, A.M., and Varner, J.E.** (1989). Plant hydroxyproline-rich glycoproteins. In *The Biochemistry of Plants*, Vol. 15, A. Marcus, ed (San Diego: Academic Press), pp. 485-520.
- Stiefel, V., Pérez-Grau, L., Albericio, F., Giralt, E., Ruiz-Avila, L., Ludevid, M.D., and Puigdomènech, P.** (1988). Molecular cloning of cDNAs encoding a putative cell wall protein from *Zea mays* and immunological identification of related polypeptides. *Plant Mol. Biol.* **11**, 483-493.
- Varner, J.E., and Lin, L.-S.** (1989). Plant cell wall architecture. *Cell* **56**, 231-239.
- Von Heijne, G.** (1981). On the hydrophobic nature of signal sequences. *Eur. J. Biochem.* **116**, 419-422.

## **Turgor-responsive gene transcription and RNA levels increase rapidly when pea shoots are wilted. Sequence and expression of three inducible genes<sup>1</sup>**

Felix D. Guerrero,<sup>1</sup> Jennifer T. Jones and John E. Mullet<sup>2</sup>

*Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX 77843–2128, USA (\*author for correspondence); <sup>1</sup> Present address: CIBA-GEIGY Biotechnology Research, Box 12257, Research Triangle Park, NC 27709–2257, USA*

Received 14 November 1989; accepted in revised form 12 March 1990

**Key words:** nuclear gene transcription, pea, plant, RNA, water deficit

### **Abstract**

Reduction of turgor in pea shoots caused the accumulation of several poly(A) RNAs. cDNA clones derived from three different poly(A) RNAs which accumulate in wilted pea shoots were isolated, sequenced and expression of the corresponding genes examined. Clone 7a encoded a 289 amino acid protein. The C-terminal 180 amino acids of this protein were homologous to soybean nodulin-26. RNA hybridizing to cDNA 7a was abundant in roots, and induced in shoots by dehydration, heat shock and to a small extent by ABA. Hydropathic plots indicate that the protein encoded by cDNA 7a contains six potential membrane spanning domains similar to proteins which form ion channels. Clone 15a encoded a 363 amino acid protein with high homology to cysteine proteases. RNA hybridizing to cDNA 15a was more abundant in roots than shoots of control plants. Dehydration of pea shoots induced cDNA 15a mRNA levels whereas heat shock or ABA treatment did not. Clone 26g encoded a 508 amino acid protein with 30% residue identity to several aldehyde dehydrogenases. RNA hybridizing to cDNA 26g was induced by dehydration of shoots but not roots and heat shock and ABA did not modulate RNA levels. Levels of the three poly(A) RNAs increased 4–6-fold by 4 h after wilting and this increase was not altered by pretreatment of shoots with cycloheximide. When wilted shoots were rehydrated, RNA hybridizing to cDNA 26g declined to pre-stress levels within 2 h. Run-on transcription experiments using nuclei from pea shoots showed that transcription of the genes which encode the three poly(A) RNAs was induced within 30 min following reduction of shoot turgor. One of the genes showed a further increase in transcription by 4 h after dehydration whereas transcription of the other 2 genes declined. These results indicate that plant cells respond to changes in cell turgor by rapidly increasing transcription of several genes. Furthermore, the expression of the turgor-responsive genes varies with respect to the time course of induction and reversibility of the wilting-induced changes.

## Introduction

Plant growth and development depend on the acquisition of water from the soil. As a consequence, exposure of plants to water-limiting environments causes a diverse set of metabolic and developmental changes. The specific plant response to water deficit depends on the severity and duration of the water deficit, the stage of development and plant species examined. For example, mild water deficit often induces stomatal closure which reduces water loss from leaves. Stomatal closure also limits gas exchange and inhibits photosynthesis. Some plants respond to this situation by switching from  $C_3$  to CAM photosynthesis which increases water use efficiency [rev. 59]. Shoot growth is also very sensitive to water deficit [3, 9, 10, 49]. Associated with inhibition of shoot growth are decreases in polyosomes and changes in poly(A) RNA populations [1, 2, 10, 41, 42]. Interestingly, shoot growth is more sensitive than root growth to water deficit [10, 64]. This is due in part to differential sensitivity of roots and shoots to abscisic acid (ABA) which increases 10- to 50-fold in water-deficient plants [10]. The mode of action of ABA is unclear but this growth regulator is known to alter  $K^+/H^+$  flux [5, 55]. Other plant growth regulators may modulate additional responses of plants to drought [rev. 67].

Cells of well watered plants maintain a positive cell turgor pressure of 0.4 to 0.8 MPa. Exposure of plants to low water potential soil or conditions where transpiration exceeds the capacity of roots to supply water to shoots often leads to loss of cell turgor. As in unicellular organisms [rev. 30], loss of plant cell turgor can lead to osmotic adjustment and the accumulation of compounds such as sucrose, proline and glycine betaine [22, 43, 65]. Sugar accumulation in hypocotyls of water-deficient soybean seedlings occurs primarily because growth decreases while sugar uptake continues [43]. Accumulation of proline and glycine betaine, however, involves increased rates of synthesis and changes in gene expression [26, 65]. Loss of turgor in plants also induces the synthesis of ABA [12, 52], which is derived from

a preexisting pool of xanthophylls such as violaxanthin [11, 19]. The induction of ABA synthesis is rapid ( $\sim 30$  min), requires nuclear gene expression [23, 54] and could involve pressure-sensitive receptors [6] or ion channels [46]. Increased levels of ABA can in turn induce changes in gene expression [4, 7, 8, 10, 18, 20, 21, 24, 27, 28, 31, 36, 40, 44, 47, 56, 57]. The expression of ABA-responsive genes is modulated during seed development [7, 8, 18, 20, 27, 31, 36, 40, 47, 57, 69], in response to plant dehydration [4, 8, 10, 20, 24, 28, 47] or low temperature [cf. 44, 56].

Recently, we reported that several pea shoot poly(A) RNAs increased in abundance when pea shoots were wilted [24]. cDNAs corresponding to several induced genes were isolated [24]. In this paper we report the sequences of three cDNAs and show that the corresponding genes are rapidly induced at the transcriptional level when the turgor of pea shoots is reduced to near zero. The expression of these genes in leaves, stems and roots was examined and induction was found to be insensitive to cycloheximide.

## Materials and methods

### *Plant growth and dehydration treatments*

Plant growth and shoot treatments were carried out as reported earlier [24]. *Pisum sativum* (Progress No. 9) was grown for 10–12 d at 23 °C in well watered vermiculite using a model E15 Conviron growth chamber. Plants were illuminated ( $70 \mu E m^{-2} s^{-1}$ ) for 18 h followed by 6 h of darkness. Plant shoots (stems plus leaves) were cut and the stem base quickly submerged under water where a portion of the lower stem was recut. Each shoot was placed in a 0.5 ml microfuge tube containing 10 mM Hepes-KOH (pH 8.0) for a 1 h hydration pretreatment with illumination at 23 °C. Shoots to be wilted were removed after pretreatment, weighed, then wilted in darkness under a stream of 23 °C air for 20 min until a 10–15% loss of fresh weight occurred. Wilted shoots were then incubated in a dark humid

chamber for 0, 0.5, 1 or 4 h prior to further analysis. One set of shoots was rehydrated by reexcision of 4 h wilted plant stems under water then submerging the stem in pretreatment buffer for 2 h in the light at 23 °C. For heat shock treatment, shoots were excised and placed in pretreatment solution for 1 h in the dark at 38 °C. Stems were kept submerged to help prevent shoot dehydration and any wilted shoots were discarded. Shoots used for examining the effects of exogenous ABA were excised as above and placed in solutions of 10 mM Hepes-KOH (pH 8.0) which contained 18  $\mu$ M ( $\pm$ ) abscisic acid for 1 h in the dark at 23 °C. Cycloheximide-treated shoots were pretreated for 1.5 h in the light at 23 °C in 10 mM Hepes-KOH (pH 8.0) which contained 20  $\mu$ g/ml cycloheximide followed by wilting and a 4 h incubation period. For one experimental condition, plants were grown hydroponically in aerated half-strength Hoagland's solution in the growth chambers described above. Water deficit was imposed on some of these plants by making the hydroponic medium 0.6 M mannitol.

#### *RNA and DNA sequence analysis*

Total RNA was isolated and analyzed by northern blotting as previously described [24] except that when utilized, LiCl precipitation of RNA was done without ethanol. Preparation of probes from selected cDNA clone inserts and the hybridization conditions have also been described [24]. Dideoxy sequencing reactions were performed on double-stranded plasmid DNA, using  $\alpha^{35}$ S-ATP (from NEN), custom-made oligonucleotide primers, and the Sequenase kit (U.S. Biochemical Corp.). DNA sequencing ladders were analyzed on 5% polyacrylamide 8.3 M urea gels and both DNA strands were sequenced to completion. DNA sequences were analyzed using the Intelligenetics computer programs (Intelligenetics, Inc., Mountain View, CA).

#### *Transcription assays of isolated nuclei*

Pea shoot nuclei were isolated using Percoll step gradients by a method similar to Hagen and Guilfoyle [25]. Approximately 10 g of pea shoots were rinsed with water followed by a 1 min soak in ice-cold diethyl ether and a quick rinse in nucleus isolation buffer (NIB; 10 mM Tris-HCl (pH 7.5), 5 mM MgCl<sub>2</sub>, 1 M sucrose, 10 mM  $\beta$ -mercaptoethanol). The remaining steps were done in a cold-room (4 °C) or on ice. All centrifugations were done at 4 °C. Shoots were quickly chopped with scissors, 75 ml NIB added and tissue homogenized using a Polytron (Brinkman PT-10 generator) for 1 min at medium high setting. The mixture was filtered through Miracloth (Calbiochem) and the filtrate centrifuged 10 min at 2500 g. The pellet was resuspended with 5 ml of NIB and further homogenized with 20 strokes in a tight-fitting Dounce homogenizer. The solution was split into two portions, layered onto separate Percoll step gradients consisting of 25, 37 and 50% Percoll in NIB, and centrifuged 15 min at 7700 g. The nuclei at the 37%–50% Percoll interface were removed, mixed with 10 ml of nuclei suspension buffer (25% v/v glycerol, 5 mM Mg acetate, 50 mM Tris-HCl (pH 8.0), 5 mM dithiothreitol, 0.1 mM EDTA) and recentrifuged for 10 min at 2500 g. The pellet of nuclei was resuspended in 2 ml of suspension buffer and recentrifuged. The resulting pellet was resuspended to a final volume of 250  $\mu$ l in suspension buffer and either used immediately for *in vitro* transcription reactions or frozen at –80 °C.

Transcription in isolated nuclei and isolation of RNA was as described by Love and Minton [37]. Each reaction containing 500  $\mu$ Ci [ $\alpha$ -<sup>32</sup>P]-UTP (800 Ci/mmol, New England Nuclear). Reaction times varied from 6–60 min depending on the experiment. Incorporation of radiolabel into RNA was determined by spotting aliquots onto Whatman 3MM paper and assaying precipitable counts by liquid scintillation counting as described in Maniatis *et al.* [39].

### Chemicals

Glycerol, agarose, guanidinium isothiocyanate, CsCl and dithiothreitol were purchased from Bethesda Research Laboratories. Percoll was obtained from Pharmacia. Nucleotide triphosphates were from Pharmacia and ultra-pure urea from Schwartz-Mann. Unless noted otherwise, all enzymes were purchased from Bethesda Research Laboratories and all other chemicals from Sigma Chemical Co.

### Results

#### *Expression of dehydration-inducible genes in intact plants*

Earlier studies [24] identified several cDNAs which correspond to genes whose poly(A) RNA levels increase when excised pea shoots are wilted and incubated in darkness. The genes' expression could have been altered due to plant excision [59, 62], loss of cell turgor, transfer of plants to darkness or to changes in plant growth regulator levels. To distinguish among the first three possibilities, pea plants were grown hydroponically in continuous light for 10 days. At this point, the medium was adjusted to 0.6 M mannitol. This treatment caused leaf wilting to occur. The plants were maintained for 4 h in the light in mannitol and then leaves were harvested for RNA extraction and northern blot analysis. The northern blot in Fig. 1 shows that mannitol treatment caused RNA hybridizing to clones 14d (a control) and 45c to decrease in abundance. Previously, RNA hybridizing to clone 45c increased when wilted pea shoots were placed in darkness (clone F in ref. 24). Other experiments showed that this gene is induced by transfer of non-wilted plants from light to darkness (data not shown). In contrast, RNA hybridizing to clones 7a, 15a and 26g (clones C, D and E in ref. 24) increased in abundance when roots of intact pea seedlings were treated with mannitol and were not affected by transfer from light to darkness. This result is consistent with previous data showing that these

genes are induced when pea leaf cell turgor is reduced to near zero [24].

#### *Expression of dehydration-inducible genes in roots, stems and leaves*

Expression of the genes corresponding to cDNAs 7a, 15a, 26g and 14d was examined in excised

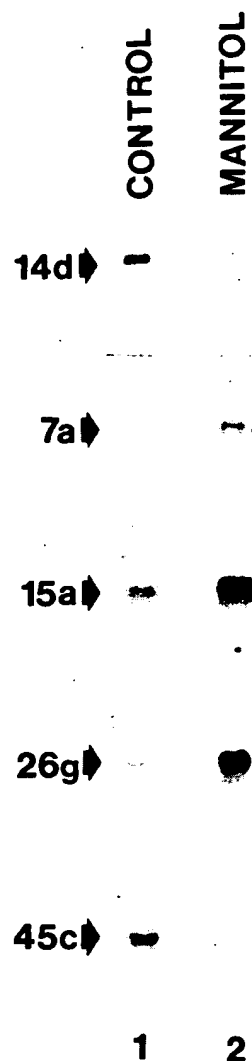


Fig. 1. Northern analysis of RNAs extracted from the shoots of control hydroponically grown pea plants (lane 1) and similar plants whose roots were exposed to 0.6 M mannitol (lane 2). Five  $\mu$ g of total RNA from control and mannitol-treated plants was separated on an agarose gel, blotted and probed using nick-translated inserts from cDNA clones 14d, 7a, 15a, 26g and 45c as described previously [24].

hydrated leaves, stems and roots and after these plant parts had been dehydrated to reduce fresh weight by 10% (leaves, stems) or exposed to 0.6 M mannitol (roots). In hydrated tissue, transcripts hybridizing to control gene 14d are most abundant in leaves and undetectable in roots (Fig. 2). The level of 14d transcripts declined in dehydrated leaves and stems. In contrast, 7a and 15a transcripts levels were most abundant in hydrated roots with lower expression in leaves. RNA levels of these genes increased in leaves and stems when these plant parts were dehydrated. A small increase in 15a and 7a RNA levels occurred when roots were treated with 0.6 M mannitol. RNA hybridizing to the gene corresponding to cDNA 26g could be detected in control roots and stems and its level increased when leaves or stems were dehydrated.

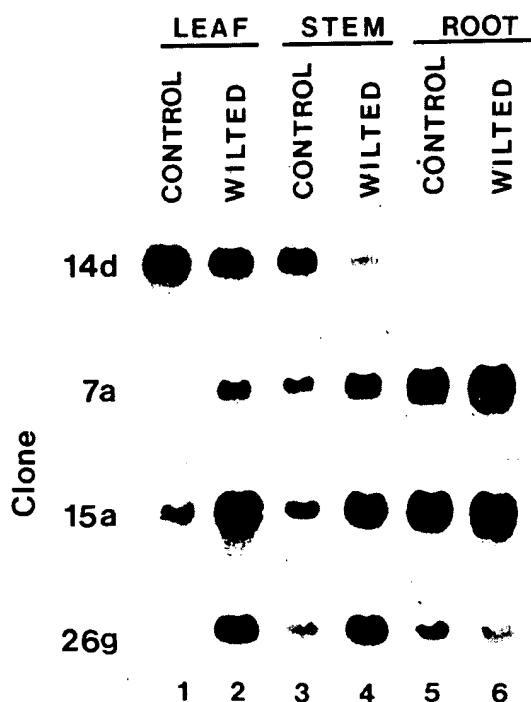


Fig. 2. Northern analysis of total RNAs from leaves (lanes 1, 2), stems (lanes 3, 4) and roots (lanes 5, 6). Leaves and stems were excised from hydrated plants and then incubated in humid boxes for 4 h (lanes 1, 3) or dehydrated to reduce fresh weight 10% and then incubated for 4 h (lanes 2, 4). Roots were excised and incubated in aerated water (lane 5) or 0.6 M mannitol for 4 h (lane 6). RNA extraction and analysis was carried out as described in Fig. 1.

*Time course of changes in RNA levels induced by wilting and influence of heat shock, ABA, rehydration and cycloheximide*

Changes in the abundance of RNAs hybridizing to cDNA clones 14d, 7a, 15a and 26g was determined as a function of time after plant dehydration, and in heat-shocked and ABA-treated plants (Fig. 3). This experiment showed that RNA hybridizing to clones 7a and 15a increased within 0.5 h after plant dehydration and that RNA levels continued to increase for at least 4 h if the reduced turgor state was maintained. RNA hybridizing to clone 26g also increased but the time course of induction was slower. We previously found that transfer of pea shoots to 38 °C for 1 h induced a large set of genes which presumably encode heat shock proteins [24]. This treatment therefore was used to test if high temperature would induce the level of RNA hybridizing to clones 7a, 15a or 26g. The results in Fig. 3 show that heat shock treatment increased the level of RNA hybridizing to clone 7a but not 15a or 26g. The ability of ABA to increase RNA levels of the dehydration-inducible genes was tested next. Excised shoots were fed 18  $\mu$ M ABA through the transpiration stream for 1 h. This treatment caused ABA levels to increase from 0.01  $\mu$ g/g fresh weight to 0.1  $\mu$ g/g fresh weight [24]. ABA accumulates to this level in wilted pea shoots within 1 h and induces changes in poly(A) RNA populations [24]. Treatment for 4 h did not increase ABA levels further, presumably because stomatal closure blocked further uptake via the transpirational stream. In the present study, treatment of seedlings for 1 h with 18  $\mu$ M ABA increased the level of RNA hybridizing to clone 7a to a small extent. RNA levels of clones 14d, 15a and 26g were not altered by ABA treatment.

The results in Fig. 3 were verified and extended by northern analysis of total RNA (Fig. 4). Total RNA was isolated from control shoots pretreated for 1 h in Hepes buffer and shoots which had been pretreated, wilted and incubated 0.5 and 4 h (Fig. 4). In addition, RNA was isolated from shoots which had been pretreated, wilted and



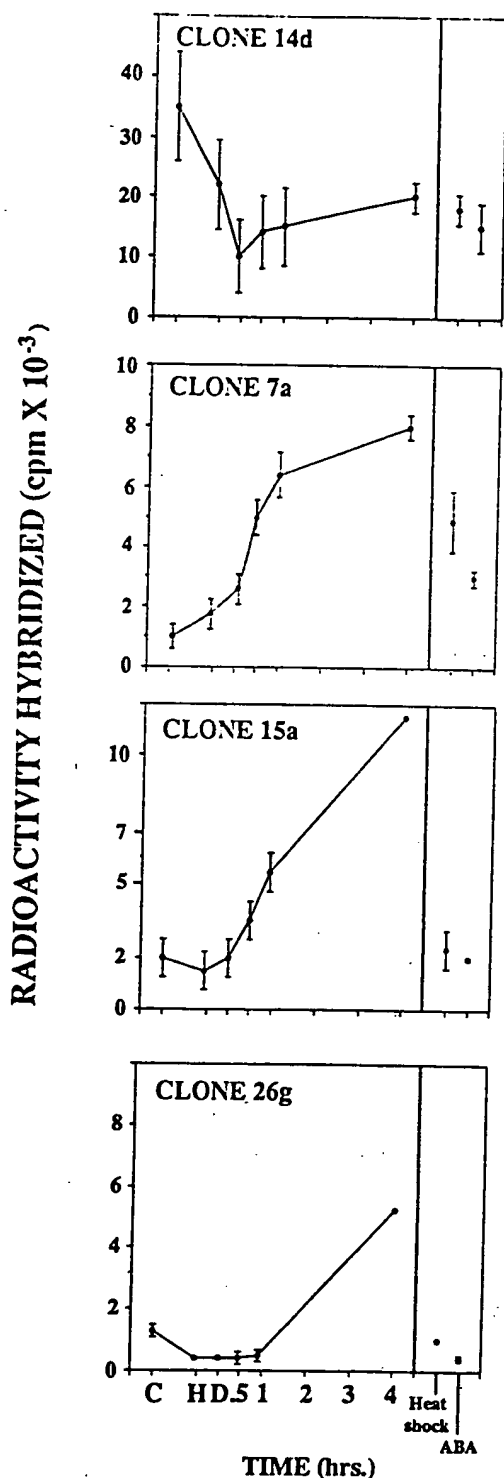


Fig. 3. Time course of changes in RNA levels induced by dehydration, heat shock or ABA. Poly(A) RNA was extracted from the shoots of control plants (C), excised shoots hydrated for 1 h (H), shoots dehydrated to produce a 10%

incubated for 4 h, and then rehydrated again by cutting the stem under water and submerging the stem in Hepes buffer for 2 h in the light at 23 °C (Fig. 4, lane 4). Cycloheximide (20 µg/ml) was fed through the transpirational stream to another set of excised shoots for 1.5 h followed by wilting and a 4 h incubation period (Fig. 4, lane 5). This treatment blocks cytoplasmic translation [48]. The northern blot in Fig. 4 confirmed the induction time course of RNAs hybridizing to clones 7a, 15a and 26g. When shoots were rehydrated for 2 h, RNA hybridizing to clone 26g declined whereas RNA hybridizing to clones 7a or 15a did not decline. This could indicate that these latter RNAs are quite stable or that rehydration does not cause a rapid decrease in transcription of these genes. This figure also shows that cycloheximide, an inhibitor of protein synthesis on 80S ribosomes, did not prevent dehydration-induced increases in RNA. This result indicates that no new proteins need to be synthesized in order to induce the levels of RNA hybridizing to clones 7a, 15a and 26g.

#### *Transcription in isolated nuclei*

Nuclei were isolated from control and 4 h wilted pea shoots and utilized for optimization studies of *in vitro* transcription reactions. To maximize the incorporation of radioactivity into RNA polymerase II transcripts, the K<sup>+</sup> optimum for RNA polymerase II in our system was determined. *In vitro* transcription reactions were done for 60 min and α-amanitin (10 µg/ml) was used to determine the contribution of RNA polymerase II to total transcription. This concentration of

weight loss (D), and dehydrated shoots incubated in a humid chamber for 0.5, 1, or 4 h. Other excised shoots were heat-shocked at 38 °C or treated with 18 µM ABA for 1 h prior to RNA extraction. Poly(A) RNA from each sample was used for cDNA probe preparation. Radiolabeled cDNA probes were hybridized to duplicate nylon filters containing 0.5 µg DNA of each gene of interest. After hybridization and washing, DNA-containing dots were excised and hybridizing radioactivity determined by scintillation counting. This experiment was done in triplicate and standard deviations are indicated for each point.

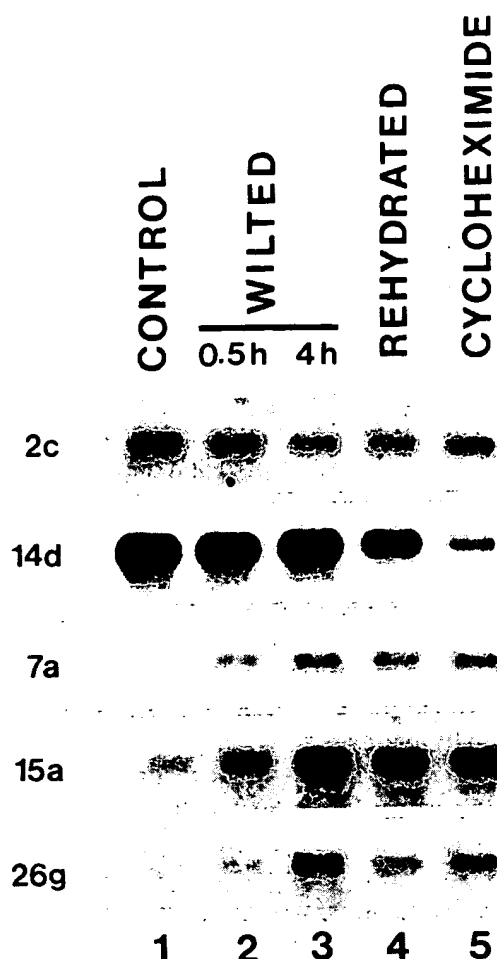


Fig. 4. Northern analysis of changes in the abundance of RNAs hybridizing to control cDNAs (2c, 14d) and to dehydration-inducible cDNAs (7a, 15a, 26g). Total RNA was extracted from hydrated pea shoots (lane 1), pea shoots dehydrated and incubated for 0.5 or 4 h (lanes 2, 3), shoots dehydrated and incubated for 4 h which were then rehydrated by incubation in buffer for 2 h (lane 4), and shoots pretreated with cycloheximide prior to dehydration and 4 h incubation (lane 5). RNA blots were prepared and probed as described in Fig. 1.

$\alpha$ -amanitin completely inhibits transcription by wheat germ RNA polymerase II [34]. These results are illustrated in Fig. 5 and show that RNA polymerase II transcription activity was maximal at 0.65 M  $K^+$ , similar to that found with *Drosophila* nuclei [37].

Transcription reaction characteristics were examined using 6 min reactions, 0.65 M  $K^+$ , and

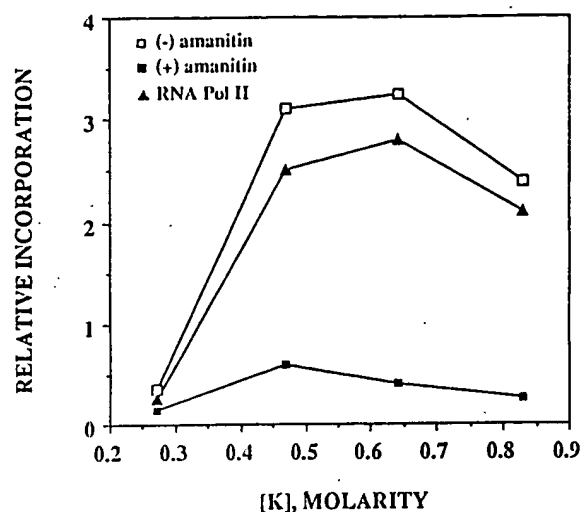


Fig. 5. Potassium optimization of transcription assays in isolated pea shoot nuclei. The upper line (open squares) denotes reactions without  $\alpha$ -amanitin, the lower line (closed squares) were reactions carried out in the presence of 10  $\mu$ g  $\alpha$ -amanitin. The line marked by closed triangles denotes  $\alpha$ -amanitin-sensitive incorporation of  $^{32}P$ -UTP into RNA in isolated nuclei.

0.65  $\mu$ Ci/ $\mu$ l [ $^{32}P$ ]-UTP (Table 1). Omission of unlabeled NTPs from the reaction solution significantly reduced the radiolabel incorporation into RNA. Including actinomycin D or  $\alpha$ -amanitin at 10  $\mu$ g/ml reduced the incorporation in both control nuclei and in nuclei from wilted shoots (Table 1). In an earlier experiment using

Table 1. Transcription in nuclei isolated from pea shoots

Sample	Percent of transcription by control nuclei <sup>1</sup>
<i>Nuclei from control shoots</i>	
Complete mix	100
- unlabeled NTPs	29
+ 10 $\mu$ g/ml $\alpha$ -amanitin	13
+ 10 $\mu$ g/ml actinomycin D	20
<i>Nuclei from 4 h wilted shoots</i>	
Complete mix	74
- unlabeled NTPs	29
+ 10 $\mu$ g/ml $\alpha$ -amanitin	16
+ 10 $\mu$ g/ml actinomycin D	33

<sup>1</sup> All reactions were carried out for 6 min. The complete transcription mix from control shoots contained  $1.34 \times 10^5$  nuclei. Incorporation over the 6 min assay was  $2.39 \times 10^4$  cpm (100% value).

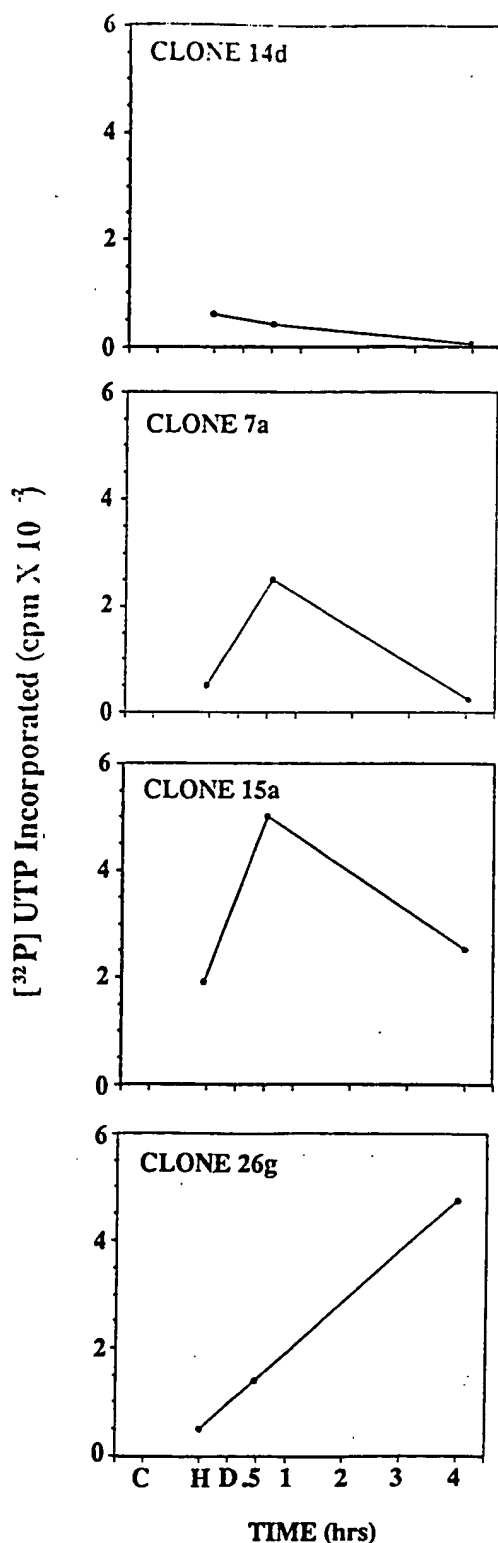


Fig. 6. Transcription of the genes corresponding to cDNAs 14d, 7a, 15a, and 26g in isolated nuclei. Pea shoots were

0.56  $\mu\text{Ci}/\mu\text{l}$  [ $^{32}\text{P}$ ]-UTP, the omission of nuclei resulted in precipitable radioactivity of 6% of the complete mix control (data not shown). The transcription activity of nuclei isolated from 4 h wilted shoots is 74% that of control shoot nuclei, indicating that wilting caused an overall inhibition of transcription activity.

#### *Transcription of turgor-responsive genes*

Transcription of the genes corresponding to cDNAs 14d, 7a, 15a and 26g was assayed in nuclei isolated from hydrated and dehydrated pea shoots. Nuclei were isolated from control shoots and shoots wilted and incubated for 0.5 and 4 h. RNA synthesized by nuclei *in vitro* was used to probe a set of replica filters containing DNA from cDNAs 14d, 7a, 15a and 26g. The results from this experiment are shown in Fig. 6. Transcription from the gene corresponding to clone 14d was not induced by wilting. Transcription of RNA hybridizing to clones 7a and 15a showed approximately a 3-fold increase by 0.5 h after shoot dehydration. However, the activity returned to approximately the pre-stress level by 4 h. In contrast, RNA levels for these genes did not decline during this period. The gene corresponding to clone 26g showed approximately a 2-fold induction of transcription activity at the 0.5 h point. Transcription activity for this latter gene continued to increase to a level approximately 6-fold over the control level by 4 h. This increase in transcription was paralleled by increases in RNA level (Fig. 4).

A second set of nuclei isolated from control shoots and shoots wilted and incubated for 0.5 and 4 h were used to repeat the experiment reported above. For these nuclei, 30 min reactions were utilized for one set of control, 0.5 and 4 h dehydrated shoots and 30 min reactions

excised at time (C), hydrated for 1 h (H) and then dehydrated until a loss of 10% fresh weight had occurred (D). Dehydrated plants were incubated for up to 4 h in a humid chamber. Nuclei were isolated at the times indicated and allowed to incorporate [ $^{32}\text{P}$ ]-UTP into RNA for 1 h. The radiolabeled RNAs were then extracted and hybridized to DNA fixed to nylon membranes.

followed by a 15 min chase with 0.5 mM unlabeled UTP for a second set of control and 4 h treated shoots. These experiments gave qualitatively similar results to those seen in Fig. 6, differing primarily in the amount of radioactivity hybridizing to the filter-bound DNA but not in the induction trends. In addition, no reduction in hybridizable RNA was evident in pulse-chase experiments and the size radiolabeled RNAs did not change significantly during the chase (data not shown).

#### DNA sequence analysis

The size of the RNAs hybridizing to the cDNAs 7a, 15a and 26g were 1.2, 1.5 and 1.8 kb respec-

tively. The cDNA inserts in clones 7a, 15a and 26g were close to the size expected for full-length copies of the mRNAs. To obtain further information concerning the products encoded by the cDNAs, DNA sequence analysis was carried out (Fig. 7A-C). Each cDNA sequence contained a single large open reading frame. The deduced amino acid sequences generated from the open reading frames are shown in Figs. 7A-C. Each deduced amino acid sequence was compared to sequences compiled in the Intelligenetics Protein Identification Resource.

Sequence analysis of cDNA 7a revealed an open reading frame encoding a putative 289 amino acid protein (Fig. 7A). The C-terminal 180 amino acids of the deduced amino acid sequence had 33% overall residue identity with soybean

AATTCGCAAGAAACCAGAACAGTCTAGAAAGAAACATGGAAGCCAAGGAACAGGATGTGTCTTTGGGAGCC	72
M E A K E Q D V S L G A	
AACAAGTTTCCAGAGAGACAGCCACTCGGGATTGCAGCTCAGAGCCAAGACGACCAAGGACTATCAAGAG	144
N K F P E R Q P L G I A A Q S Q D E P K D Y Q E	
CCACCACGGGGCCACTTTTGTAGCCCTCGGAGCTGACTTCATGGTCTTTTACAGAGCCGGATAGCCGAG	216
P P P A P L F E P S E L T S W S F Y R A G I A E	
TTATAGCCACTTTTCTTTTCTTTATATAACCGTTTAAACCGTCATGGGCGTTGTTGAGAAAGTTCCAAG	288
F I A T F L F L Y I T V L T V M G V V R E S S K	
TGCAAAACGGTTGGTATTCAAGGAATCGCTTGGGCTTTTGGTGGCATGATAATTGCTCTCGTTTACTGCACC	360
C K T V G I Q G I A W A F G G M I F A L V Y C T	
GCTGGAATCTCAGGGGTCATATAAACCCAGCTGTGACGTTTGGGCTATTTTGGCGAGAAATNGTCGTTG	432
A G I S G G H I N P A V T F G L F L A R K S L	
ACAAGAGCAATATTCTACATGGTGAATGCAAGTGTGGGTGCTATATGCGGTGCTGGTGGTGAAGGTTTT	504
T R A I F Y M V M Q V L G A I C G A G V V K G F	
GAAGGTAAACAAAGTTTGGAGATCTGAACGGTGGTCCCAACTTTTGTGCTCTGGTTACACCAAAGGTGAT	576
E G K Q R F G D L N G G A N F V A P G Y T K G D	
GGACTTGGTGTGAAATTTGTGGCACTTTTCTTTGTATACAGTTTTCTCAGCCACTGATGCTAAACGT	648
G L G A E I V G T F I L V Y T V F S A T D A K R	
AGCGCCAGAGACTCTCATGTTCTTATTTGGCACCGTTGCCAATTGGATTGCGTGTGTTCTTGGTGCATTGG	720
S A R D S H V P I L A P L P I G F A V F L V H L	
GCACTATACCAATTACTGGAAGTGTATTAAACCTGCCAGAGTCTTGGTCTGCTATTGTCTTCAACAAG	792
A T I P I T G T G I N P A R S L G A A I V F N K	
AAAATGGTTGGAATGATCACTGGATTTCCTGGGTTGGACATTTATTGGAGCAGCTCTTGCAGCACTATAC	864
K I G W N D H W I F W V G P F I G A A L A A L Y	
CATCAAGTTGTTATCAGAGCCATTCCCTTCAAGTCTAAGTGATTGAATTGAATCAACCGTTCTTGATCAAT	936
H Q V V I R A I P F K S K	
CATATCATGTTTGTGGATTTCATTTTCATCATGAAATGAATAATCITTATGTATGATATTTTTTCTGAATTT	1008
TTAAATGTTTGTGTAATTTGTATGTAAAGAAGTACTATTTATTATGATTATTAGGTGTAGCATGTGTGT	1080
GTTGCGTAATGGGTCAGCTCTCGTGTGTAATGTGATTGTCGCAATCTAAAATAACACTCGAGTCATCA	1152
TTTCTGAAAAAAAGG	1168

Fig. 7A.

Fig. 7. DNA sequences of cDNA 7a (Fig. 7A, this page), cDNA 15a (Fig. 7B) and cDNA 26g (Fig. 7C). Deduced amino acid sequences are shown below the corresponding DNA codons. Amino acid regions which have sequence similarities to other proteins are underlined and discussed in the text.

AATTCCGCTCTTCTCACCACGAACCTTCGCTCCCCACCATGGATCGCGTTTCTCTTCGCTCTCTTC	72
M D R R F L F A L F L	
TCTTCCCGCCGTAGCAACCGCGTAACCGACGACACCAACACGACGACTTCATAATCCGTCAAGTCGTCG	144
F A A V A T A V T D D T N N D D F I I R Q V V D	
ATAACGAAGAAGATCACTTGTAAACGAGAGCATCACTTCACTTCAAGTCAAAGTTCAGCAAAAGCT	216
N E E D H L L N A E H H F T S F K S K F S K S Y	
ACGCAACAAAAGAGGAACACGATTACCGATTGCGCGTGTCAAATCGAATCTAATCAAAGCGAAGCTCCACC	288
A T K E E H D Y R F G V F K S N L I K A K L H Q	
AGAATCGCGATCCTACTGCGGAGCAGGAATACTAAGTCTCCGATTAACTGCGTCAGAGTTTCGCGGTC	360
N R D P T A E H G I T K F S D L T A S E F R R Q	
AGTTTCTCGGACTTAAGAAACGTCCTCGTTTGCCTGCTCATGCTCAGAAGGCTCCGATCCTCCCTACCACCA	432
F L G L K K R L R L P A H A Q K A P I L P T T N	
ACCTCCCGGAGGACTTCGATTGCGGTGAGAAAGCGCGTTACTCCCGTCAAGGACCAGGGTTTCATGCGGTT	504
L P E D F D W R E K G A V T P V K D Q G S C G S	
CGTGTGCGGCTTTAGCACAAACCGAGCTTTAGAAGGAGCACACTATCTGGCTACTGGGAAGCTTGTAAAGCC	576
C W A F S T T G A L E G A H Y L A T G K L V S L	
TTAGCGAACAACAGCTTGTGGATTGTGATCATGTGTGATCCAGAGCAAGCTGGGTTCATGTGACTCGGGCT	648
S E Q Q L V D C D H V C D P E Q A G S C D S G C	
GTAATGGTGGATTGATGAACAATGCTTTGCAATATTTACTCGAGTCTGGGGAGTAGTGCAAGAAAAGGACT	720
N G G L M N N A F E Y L L E S G G V V Q E K D Y	
ATGCTTACACCGGAAGAGATGGCTCCTGCAAAATTTGACAAAAGCAAAGTTGTTGCTTCCGTATCTAATTTCA	792
A Y T G R D G S C K F D K S K V V A S V S N F S	
GTGTGGTTACCTAGACGAAGACCAAAATGCTGCAATCTAGTAAAGAAATGGCCCTCTTGCAAGTTGCTATTA	864
V V T L D E D Q I A A N L V K N G P L A V A I N	
ATGCAGCTGGATGCAGACATACATGAGTGGCGTCTCATGCCCATATGCTGTGCCAAATCGCGCTGGATC	936
A A W M Q T Y M S G V S C P Y V C A K S R L D H	
ATGGTGTCTCTAGTTGGTTTGGAAAAGGTGCTTATGCTCCATTGCAAGTGAAGAAAAGCCTTACTGGA	1008
G V L L V G F G K G A Y A P I R L K E K P Y W I	
TCATTAAGAACTCTGGGGGAGAAATGGGAGAGCAGGGATATTACAAGATCTGCAGAGGTAGAAATGTAT	1080
I K N S W G Q N W G E Q G Y Y K I C R G R N V C	
GTGGAGTCGATTCAATGGTTTCACTGTAGCTGCAGCTCAATCCAACATTAAATACAGGGATGCTGATTCT	1152
G V D S M V S T V A A A Q S N H	
AGTCTGCTCTCATCTCATGTGACTCTTAAGTTAGTTGAATTTGTGTAATATATTATCATGAAGTTGTGAAT	1224
GTATCTTCTGGITATGCAAAAATATCTCTAAGTAGGATCAAGAGACTGCCACGATCATGTTATTTAAGTA	1296
GTTGGTTATGCTTGTCTGTACTGTGGTGATGTTTATCCTTAGCCATGCCAGCATAGCAGGTAAATGAAGTG	1368
GAAAAATAATTATAAACTTTATTATTTATGAGTTTATGGAACAACATATGTAATATATATTCTATTAGGTTA	1440
AAAAAAGGAATT	1453

Fig. 7B.

nodulin-26 [17]. High amino acid sequence similarities were detected between amino acids 112 to 122 (10 of 11 identical) and 239 to 250 (9 of 11 identical) (regions underlined in Fig. 7A). Hydrophobic plots revealed 6 stretches of hydrophobic amino acids which could correspond to membrane-spanning domains (Fig. 8A).

The protein encoded by cDNA 15a had 41% overall residue identity when compared to cysteine proteinase 1 from *Dictyostelium* (Fig. 7B). Amino acids 143 to 190 showed 76% sequence identity to the cysteine protease. The

protein encoded by cDNA 15a also showed regions of 50 to 80% sequence identity with other thiol proteases (cathepsin H, papain, and aleurain) and contained conserved active site residues (Cys 154; His 300) [60, 66, 68]. Based on this analysis we tentatively identify the protein encoded by cDNA 15a as a thiol protease. A hydrophobic plot of this protein revealed 2 long stretches of hydrophobic amino acids (Fig. 8B). The first is a 16 amino acid region immediately adjacent to a small basic N-terminal domain. Signal sequences in secreted proteins often show

AATTTCGCCTAATCAACAACACTTCGTCAATATCCCATATAAAGCCTCTCAATTATTCAATTCATTCAACTTCACTT	72
CAGTITCCAGAGAAATGGGTTCTGATAGCAACAATTGGGATTCTTGAAGGAGATCGGTTTGGGTGCTACTA M G S D S N N L G F L K E I G L G A T N	144
ACATCGGTTCTTTTCATCAATGGCCAATGGAAAGCCAAACGGTCTACTGTTTCATTCGCTCAATCCTTCCACCA I G S F I N G Q W K A N G P T V H S V N P S T N	216
ATCAGGTTATTGCTTCCTGACTGAAGCAACTTTGGATGATTATGAAGAGGGATTGAGAGCTAGCAGTGAAG Q V I A S V T E A T L D D Y E E G L R A S S E A	288
CAGCAAGACATGGAGGACTGTTCCCGCACCCGAAAAGAGGTGAGATTGTAAGACAGATTGGGGATCGGTTG A K T W R T V P P A K R G E I V R Q I G D A L R	360
GGGCTAAGTTGGATCCACTTGGTAGGTTGGTGGCTCTTGAGATGGGAAAAATTTGCTGAAGGAATTTGGTG A K L D P L G R L V A L E M G K I L A E G I G E	432
AGGTTTCAGGAAATTTATGATATGTGTGATTATTCTGTTGGGCTGAGCCGACAACTGAATGGATCAATTATCC V Q E I I D M C D Y S V G L S R Q L N G S I I P	504
CATCAGAACGTCAGAGCATATGATGTTTGAGGTATGGAACCCATTAGGAATCGTTGGTGTAATCACTGCCT S E R P E H M M F E V W N P L G I V G V I T A F	576
TCAACTTTCAGTGTGCTGACTAGGATGGAATGCTTGCAATTGCACTAGTCGGTGGTAACACTGTTGTTTGA N F P C A V L G W N A C I A L V R G I N T V D W K	648
AGGGCGCGCCAACTACTCCATTGATAACTGTGCTGTGACAAGCTGATGCTGAAGTTTTCAGAGGAACA G A P T T P L I T V A V T K L I A E V F E R N N	720
ATTTACCCGAGCAATATTTACCGCTCTCTGTGGGGTGTGATATTGGGCACGCAATAGCGAAAGACACAC L P G A I F T A L C G G A D I G H A I A K D T R	792
GCATTCTCTTGGTTTCATTACTGGAAGCTCAAAGGTGGGCGCGCTGGTCCAGCAAACAGTGAGTCAAAGAT I P L V S F T G S S K V G A A L V Q Q T V S Q R F	864
TTGGCAAACCTTGGTTGAGTTGAGTGGTAACAATGCAATAATAGTCATGGATGATCGGCATACACATTGG G K T L L E L S G N N A I I V M D D A D I T L A	936
CTGTGCGCTCTATTCTTTGTCAGTGTGGTACTGCTGGTCAGCGTTGTACAACCTGTCTGAGACTGTATC V R S I T F F A C A G T G C T A G Q R C T T C R R L Y L	1008
TGCATGAGAGTGTATTATGCAAAACGTGCTTGAACAACCTTACCGCACTCTACAAACAAGTCAAAATTTGGGAATC H E S V Y A N V L E Q Q L T A L Y K Q V K I G N P	1080
CTCTGGAGGAAGGGACACTAGTTGGGCTCTGTCATACTCGTTCTCGCGTGGAAAATTTTAAGAATGGTATTT L E E G T L V G P L H T R S A V E N F K N G I S	1152
CTGCAATTAATCTCAGGAGGGAAGATTGTAACAGGGGGTCTGTATTAGAGTCCGAAGGAAACTTTTGTC A I K S Q G K I V T G G S V L E S T G E N F V	1224
TGCCAACGATTGTGAGATTCTGCAGATGCTGCTGTAGTTAAAGAAGAATTGTTGCTCCAGTTCTCTATG P T I V E I S A D A A V V K E E L F A P V L Y V	1296
TTATGAAATTAAGGATCTTGAAGAAGCAATTGCCTTGAACAATTCGTAACCTCAAGGATTGAGTAGTTCCA M K F K D L E E A I A L N N S V P Q G L S S S I	1368
TCTTCACACAAAAGCCTTCTACTATATTCAAATGGATAGGGCCAAGTGAAGTGATTGTGGTATTGTAATG F T Q K P S T I F K W I G P S G S D C G I V N V	1440
TGAACATACCAACAAATGGAGCTGAGATTGGAGGTGCCTTTGGTGGAGAAAAGGCAACTGGTGGTGGTGGT N I P T A A E I G A F F G G E K A T G G G R E	1512
AAGCTGGAAGTGACTCGTGAAGCAATACATGCGCGGTTCTACATGTACCAATTAATGGAAGTAATTTG A G S D S W K Q Y M R R S T C T I N Y G S E L P	1584
CATTAGCTCAGGGAATTAACTTTGGCTAATTTTGAAGTTTGGTAGACATGAGTTCAAATGGTGGAAACACTC L A Q G I N F G	1656
TTTTCAAATGGTTGTTTCAACAGATATGTTGTGAAAGTGGCACGGTTGATGTTTGAGATTGAATAACAAA	1728
TGCCCAATTAATTGAGTGTGG	1751

**Fig. 7C**

a similar arrangement [29]. It has been reported that aleurain, a thiol protease in barley, is secreted into vacuoles [58].

The sequence of cDNA 26g revealed an open reading frame which encoded a 508 amino acid

protein (Fig. 7C). The deduced amino acid sequence had 30% residue identity with an aldehyde dehydrogenase (*aldDH*) from *Aspergillus nidulans* [51]. Regions of higher sequence identity were scattered throughout the open reading frame

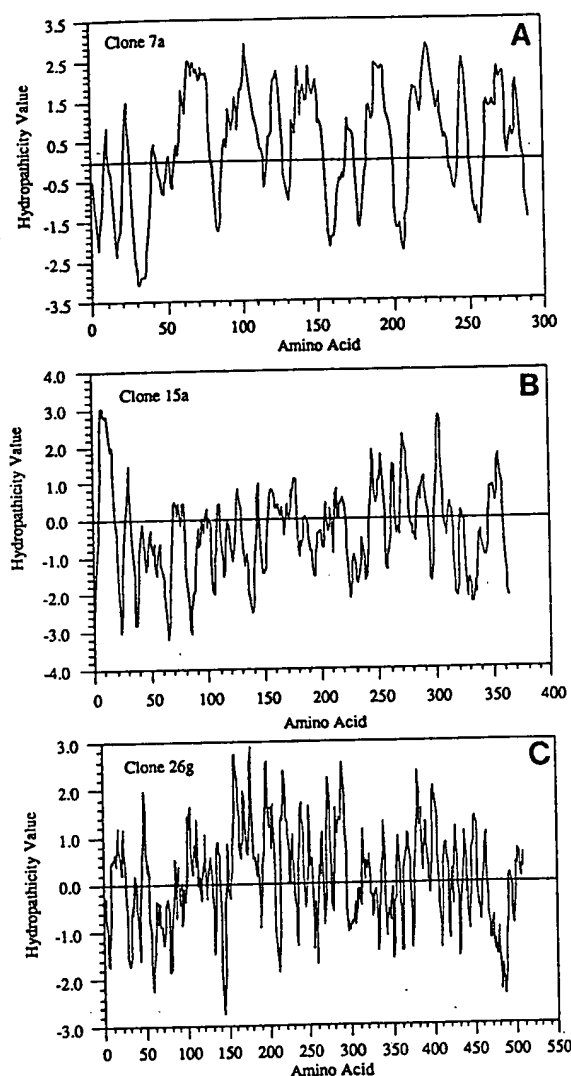


Fig. 8. Hydropathy plots of the putative proteins encoded by cDNA 7a (A), cDNA 15a (B) and cDNA 26g (C). Positive values indicate hydrophobic regions. Plots were generated by the method of Kyte and Doolittle [35] using 6 amino acid intervals for comparison.

(i.e., a.a. 46–55, 7/9; a.a. 183–196, 9/13; a.a. 406–415, 7/9; a.a. 465–480, 10/15 identical amino acids) (regions underlined in Fig. 7C). Aldehyde dehydrogenases have a highly conserved sequence (VTLELGKKS) followed by a cysteine 28 amino acids downstream of this sequence [53]. The protein encoded by cDNA 26g only had poor conservation of this sequence (TLLESGNN) (underlined with an arrow in Fig. 7C) but did retain a cysteine (Cys 301) 29

amino acids from this amino acid sequence. These data suggest that the deduced 508 amino acid protein is related to aldehyde dehydrogenases but has diverged in significant ways from this group of enzymes. Hydropathic plots of this protein revealed a hydrophobic domain of 21 amino acids near the protein's N-terminus and a 36 amino acid hydrophobic domain located between amino acid 152 and 188 (Fig. 8C).

## Discussion

Exposure of plants to water deficit often results in inhibition of shoot growth, decreased photosynthesis, osmotic adjustment and altered plant growth regulator levels [rev. 33, 67]. Some of these responses may be the direct consequence of reduced plant cell turgor. However, other responses involve changes in gene expression. Several authors have reported that water deficit causes a decrease in polysome content in growing tissues [1, 10, 32, 41, 42, 45, 63]. Decreases in polysome content and protein synthesis activity probably are global responses to reduced growth rates. In addition, specific changes in poly(A) RNA populations and protein synthesis are observed in plants exposed to low water potentials [2, 13, 14, 15, 24, 42, 45, 50]. Water deficit-induced changes in the abundance of selected transcripts and proteins suggest that specific biochemical responses of plants to dehydration involve changes in gene expression. For example, there is a correlation between seed desiccation, ABA accumulation and increased expression of specific genes [rev. 16]. The proteins encoded by this set of genes are hydrophilic, have repeating domains [8, rev. 16] and can be glycine-rich [21]. RNA hybridizing to these genes is not abundant in non-stressed plants but can be induced by dehydration [i.e., 8, 47]. The proteins encoded by these ABA-responsive genes have been proposed to protect cell structures from dehydration-induced damage [rev. 16]. In addition to the ABA-responsive genes discussed above, other genes are induced by water deficit which are ABA-insensitive. For example, in-

creases in ABA which occur in wilted plants are dependent on nuclear gene transcription [17, 41] and several changes in poly(A) RNA population were observed prior to accumulation of ABA in wilted pea leaves [24]. Likewise, soybean seedlings exposed to mild water deficit ( $-0.3$  MPa vermiculite at 100% relative humidity) exhibited changes in poly(A) RNA populations which could not be induced by exogenous ABA [10].

In this paper, the expression of 3 genes which show increased transcription and RNA levels in wilted pea shoots were characterized. ABA caused a small increase in RNA hybridizing to cDNA 7a but not to cDNAs 15a or 26g. These latter genes might respond to higher levels of ABA or to extended ABA treatment. However, the lack of ABA-responsiveness of these genes is not surprising because in a previous study, most of the poly(A) RNAs induced rapidly in response to shoot dehydration were not induced by ABA [24]. Furthermore, transcription of the genes corresponding to cDNAs 7a, 15a and 26g was induced within 30 min following shoot dehydration. However, increases in ABA in dehydrated pea shoots could only be detected after 30 min dehydration to zero turgor [24]. Finally, cycloheximide inhibits induction of ABA in wilted pea shoots [23] but this treatment did not prevent the increases in RNA hybridizing to cDNAs 7a, 15a and 26g. It was possible that the genes induced by plant dehydration would also be responsive to heat shock. In fact, one of these genes (corresponding to cDNA 7a) was induced when plants were exposed to  $38^{\circ}\text{C}$  temperatures. Genes corresponding to the other cDNAs (15a, 26g) were not induced by this treatment. This is consistent with previous studies showing that most of the changes in poly(A) RNA populations induced by dehydration of plants to zero turgor were not induced by high temperature [24]. More severe plant dehydration, however, may induce some heat shock genes [28].

Transcription assays using isolated nuclei showed that transcription of the genes corresponding to cDNA clones 7a, 15a and 26g is increased within 30 min of plant dehydration.

Transcription of RNA hybridizing to 7a and 15a declined between 30 min and 4 h after plant dehydration. However, the corresponding *in vivo* RNA levels of these genes did not decline suggesting that the RNAs are relatively stable. RNAs hybridizing to clones 7a and 15a also remain high when dehydrated seedlings are rehydrated for 2 h. In contrast, RNA hybridizing to clone 26g declined when pea shoots were rehydrated. The increase in RNA hybridizing to clones 7a, 15a and 26g was not prevented by pretreatment of pea shoots with cycloheximide. This indicates that the increase in RNA abundance is not mediated by proteins synthesized in response to dehydration.

Sequence analysis of the 3 cDNAs revealed that cDNA 7a encodes a protein with some sequence similarity to soybean nodulin-26, cDNA 15a encodes a putative thiol protease, and cDNA 26g a protein with some homology to aldehyde dehydrogenases. RNA encoding a thiol protease was reported to increase tomato in response to low temperature [60]. Shaffer and Fisher [60] noted that proteases could alter metabolism by increasing protein turnover rates or proteolytically activating specific proteins. Another function which might be common to low temperature and dehydration involves degradation of polypeptides denatured due to the stress. In addition, the thiol protease could be involved in degrading vegetative storage proteins located in vacuoles. The mobilized amino acids would then be available for the synthesis of new proteins in response to stress or for osmotic adjustment. The homology between the deduced amino acid sequence of cDNA 26g and aldehyde dehydrogenases was interesting because betaine aldehyde dehydrogenase and ABA aldehyde dehydrogenase are involved in responses of some plants to water deficit. However, pea plants do not accumulate significant levels of betaine [70] and a highly conserved amino acid sequence found in aldehyde dehydrogenases [53] was not highly conserved in the deduced amino acid sequence of cDNA 26g. Finally, the amino acid sequence of cDNA 7a showed some similarity to soybean nodulin-26 which is a peribacterioid membrane protein [17]. The protein encoded by cDNA 7a



had six potential membrane-spanning domains. This number of membrane-spanning domains is common to proteins which form  $K^+$ ,  $Na^+$  or  $Ca^{2+}$  channels [38]. The induction of ion channels in response to reduced cell turgor would not be unexpected [30]. The high expression of this gene in roots vs shoots of control plants may also provide some clue to this gene function in well watered seedlings.

### Acknowledgements

The authors thank Mrs. Sharyll Pressley for her assistance in preparing this manuscript.

This research was supported by Grant No. 8801234 from the United States Department of Agriculture to J.E.M.

### References

1. Bewley JD, Larsen KM: Differences in the responses to water stress of growing and non-growing regions of maize mesocotyls: Protein synthesis on total, free and membrane-bound polyribosome fractions. *J Expt Bot* 33: 406–415 (1982).
2. Bewley JD, Larson KM, Papp JET: Water-stress-induced changes in the pattern of protein synthesis in maize seedling mesocotyls: A comparison with the effects of heat shock. *J Expt Bot* 34: 1126–1133 (1983).
3. Boyer JS: Relationship of water potential to growth of leaves. *Plant Physiol* 43: 1056–1062 (1968).
4. Bray EA: Drought- and ABA-induced changes in polypeptide and mRNA accumulation in tomato leaves. *Plant Physiol* 88: 1210–1214 (1988).
5. Brown AW:  $H^+$  excretion in mechanically isolated mesophyll cell from *Asparagus* cladophylls. *Plant Sci Lett* 22: 155–160 (1981).
6. Chiang RG, Davey KG: A novel receptor capable of monitoring applied pressure in the abdomen of an insect. *Science* 241: 1665–1667 (1988).
7. Choi JH, Liu L, Borkird C, Sung ZR: Cloning of genes developmentally regulated during plant embryogenesis. *Proc Natl Acad Sci USA* 84: 1906–1910 (1987).
8. Close TH, Kortt AA, Chandler PM: A cDNA-based comparison of dehydration-induced proteins (dehydrins) in barley and corn. *Plant Mol Biol* 13: 95–108 (1989).
9. Cosgrove D: Biophysical control of plant cell growth. *Ann Rev Plant Physiol* 37: 377–405 (1986).
10. Creelman RA, Mason HS, Bensen RJ, Boyer JS, Mullet JE: Water deficit and ABA cause differential inhibition of shoot vs root growth in soybean seedlings. Analysis of growth, sugar accumulation and gene expression. *Plant Physiol* (in press).
11. Creelman RA, Zeevaart JAD: Incorporation of oxygen into abscisic acid and phaseic acid from molecular oxygen. *Plant Physiol* 75: 166–169 (1984).
12. Creelman RA, Zeevaart JAD: Abscisic acid accumulation in spinach leaf slices in the presence of penetrating and nonpenetrating solutes. *Plant Physiol* 77: 25–28 (1985).
13. Damerval C, Vartanian N, de Vinne D: Differential two-dimensional protein patterns as related to tissue specificity and water conditions in *Brassica napus* var *oleifera* root system. *Plant Physiol* 86: 1304–1309 (1988).
14. Dasgupta J, Bewley JD: Desiccation of axes of *Phaseolus vulgaris* during development of a switch from a development pattern of protein synthesis to a germination pattern. *Plant Physiol* 70: 1224–1227 (1982).
15. Dhindsa RS, Cleland RE: Water stress and protein synthesis. I. Differential inhibition of protein synthesis. *Plant Physiol* 55: 778–781 (1975).
16. Dure L, Crouch M, Harada J, Ho THD, Mundy J, Quatrano R, Thomas T, Sung ZR: Common amino acid sequence domains among the LEA proteins of higher plants. *Plant Mol Biol* 12: 475–486 (1989).
17. Fortin MG, Morrison NA, Verma DPS: Nodulin-26, a peribacteroid membrane nodulin is expressed independently of the development of the peribacteroid compartment. *Nucleic Acids Res* 15: 813–824 (1987).
18. Galau GA, Bijaisoradat N, Hughes DW: Accumulation kinetics of cotton late embryogenesis-abundant mRNAs and storage protein mRNAs: Coordinate regulation during embryogenesis of the role of abscisic acid. *Devel Biol* 123: 198–212 (1987).
19. Gamble PE, Mullet JE: Inhibition of carotenoid accumulation and abscisic acid biosynthesis in fluridone-treated dark-grown barley. *Eur J Biochem* 160: 117–121 (1986).
20. Goday A, Sánchez-Martínez D, Gómez J, Puigdomènech P, Pagès M: Gene expression in developing *Zea mays* embryos: Regulation by abscisic acid of a highly phosphorylated 23- to 25-kD group of proteins. *Plant Physiol* 88: 564–569 (1988).
21. Gómez J, Sánchez-Martínez D, Stiefel V, Rigau J, Puigdomènech P, Pagès M: A gene induced by the plant hormone abscisic acid in response to water stress encodes a glycine-rich protein. *Nature* 334: 262–264 (1988).
22. Grumet R, Hanson AD: Genetic evidence for an osmoregulatory function of glycinebetaine accumulation in barley. *Aust J Plant Physiol* 13: 353–364 (1986).
23. Guerrero F, Mullet JE: Increased abscisic acid biosynthesis during plant dehydration requires transcription. *Plant Physiol* 80: 588–591 (1986).
24. Guerrero FD, Mullet JE: Reduction of turgor induces rapid changes in leaf translatable RNA. *Plant Physiol* 88: 401–408 (1988).
25. Hagen G, Guilfoyle TJ: Rapid induction of selective transcription by auxins. *Mol Cell Biol* 5: 1197–1203 (1985).

26. Hanson AD, Scott NA: Betaine synthesis from radioactive precursors in attached, water-stressed barley leaves. *Plant Physiol* 66: 342-348 (1980).
27. Harada J, DeLisle A, Baden C, Crouch M: Unusual sequence of an abscisic acid-inducible mRNA which accumulates late *Brassica napus* development. *Plant Mol Biol* 12: 395-401 (1989).
28. Heikkilä JJ, Papp JET, Schultz GA, Bewley JD: Induction of heat shock protein messenger RNA in maize mesocotyls by water stress, abscisic acid, and wounding. *Plant Physiol* 76: 270-274 (1984).
29. Heijne GV: A new method for predicting signal sequence cleavage sites. *Nucl Acids Res* 14: 4683-4690 (1986).
30. Higgins CF, Cairney J, Stirling DA, Sutherland L, Booth IR: Osmotic regulation of gene expression: Ionic strength as an intracellular signal? *Trends Biochem Sci* 12: 239-244 (1987).
31. Hong B, Uknes SJ, Ho THD: Cloning and characterization of a cDNA encoding a mRNA rapidly-induced by ABA in barley aleurone layers. *Plant Mol Biol* 11: 495-506 (1988).
32. Hsiao TC: Rapid changes in levels of polyribosomes in *Zea mays* in response to water stress. *Plant Physiol* 46: 281-285 (1970).
33. Hsiao TC: Plant responses to water stress. *Ann Rev Plant Physiol* 24: 519-570 (1973).
34. Jendriak JJ, Guilfoyle TJ: Eukaryotic RNA polymerases: Comparative subunit structures, immunological properties and  $\alpha$ -amanitin sensitivities of the class II enzymes from higher plants. *Biochemistry* 17: 1322-1327 (1978).
35. Kyte J, Doolittle RF: A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 157: 105-132 (1982).
36. Lin L-S, Ho T-HD: Mode of action of abscisic acid in barley aleurone layers. Induction of new proteins by abscisic acid. *Plant Physiol* 82: 289-297 (1986).
37. Love JD, Minton KW: Screening of  $\lambda$  library for differentially expressed genes using *in vitro* transcripts. *Anal Biochem* 150: 429-441 (1985).
38. Maelicke A: Structural similarities between ion channel proteins. *Trends Biochem Sci* 13: 199-202 (1988).
39. Maniatis T, Fritsch EF, Sambrook J: *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York (1982).
40. Marcotte WR, Bayley CC, Quatrano RS: Regulation of a wheat promoter by abscisic acid in rice protoplasts. *Nature* 335: 454-457 (1988).
41. Mason HS, Matsuda K: Polyribosome metabolism, growth and water status in the growing tissues of osmotically stressed plant seedlings. *Physiol Plant* 64: 95-104 (1985).
42. Mason HS, Mullet JE, Boyer JS: Polysomes, messenger RNA, and growth in soybean stems during development and water deficit. *Plant Physiol* 86: 725-733 (1988).
43. Meyer RF, Boyer JS: Osmoregulation, solute distribution, and growth in soybean seedlings having low water potentials. *Planta* 151: 482-489 (1981).
44. Mahapatra SS, Poole RJ, Dhindsa RS: Absciscic acid-regulated gene expression in relation to freezing tolerance in alfalfa. *Plant Physiol* 87: 468-473 (1988).
45. Morilla CA, Boyer JS, Hageman RH: Nitrate reductase activity and polyribosomal content of corn (*Zea mays* L.) having low leaf water potentials. *Plant Physiol* 51: 817-824 (1973).
46. Morris CE, Sigurdson WJ: Stretch-inactivated ion channels coexist with stretch-activated ion channels. *Science* 243: 807-809 (1989).
47. Mundy J, Chua N-H: Absciscic acid and water-stress induce the expression of a novel rice gene. *EMBO J* 7: 2279-2286 (1988).
48. Mullet JE, Klein RR, Grossman AR: Optimization of protein synthesis in isolated higher plant chloroplasts. Identification of paused translation intermediates. *Eur J Biochem* 155: 331-338 (1986).
49. Nonami H, Boyer JS: Turgor and growth at low water potentials. *Plant Physiol* 89: 798-804 (1989).
50. Oliver MJ, Bewley JD: Plant desiccation and protein synthesis. VI. Changes in protein synthesis elicited by desiccation of the moss *Tortula ruralis* are effected at the translational level. *Plant Physiol* 74: 923-927 (1984).
51. Pickett M, Gwynne DI, Buxton FP, Elliott R, Davies RW, Lockington RA, Scazzocchio C, Sealy-Lewis HM: Cloning and characterization of the *aldA* gene of *Aspergillus nidulans*. *Gene* 51: 217-226 (1987).
52. Pierce M, Raschke K: Synthesis and metabolism of abscisic acid in detached leaves of *Phaseolus vulgaris* L. after loss and recovery of turgor. *Planta* 153: 156-165 (1981).
53. Pietruszko R: *Biochemistry and Physiology of Substance Abuse* 1: 89-127 (1989).
54. Quarrie SA, Lister PG: Effects of inhibitors of protein synthesis on abscisic acid accumulation in wheat. *Z Pflanzenphysiol* 114: 309-314 (1984).
55. Reed NR, Bonner BA: The effect of abscisic acid on the uptake of potassium and chloride into *Avena* coleoptile sections. *Planta (Berl)* 116: 173-185 (1974).
56. Robertson AJ, Gusta LV, Reaney MJT, Ishikawa M: Identification of proteins correlated with increased freezing tolerance in brome grass (*Bromus inermis* Leyss. cv Manchar) cell cultures. *Plant Physiol* 86: 344-347 (1988).
57. Rodriguez D, Nicolás G, Aldasoro JJ, Hernández-Nistal J, Babiano MJ, Matilla A: Altered development of polysomal RNA activity in chick-pea (*Cicer arietinum* L.) embryonic axes. Effects of abscisic acid and temperature. *Planta* 164: 517-523 (1985).
58. Rogers JC, Dean D, Heck GR: Aleurain: A barley thiol protease closely related to mammalian cathepsin H. *Proc Natl Acad Sci USA* 82: 6512-6516 (1985).

59. Ryan CA, An G: Molecular biology of wound-inducible proteinase inhibitors in plants. *Plant Cell Envir* 11: 345-349 (1988).
60. Schaffer MA, Fischer RL: Analysis of mRNAs that accumulate in response to low temperature identifies a thiol protease gene in tomato. *Plant Physiol* 87: 431-436 (1988).
61. Schmitt JM, Michalowski C, Bohnert HJ: Gene expression during CAM induction under salt stress in *Mesembryanthemum*: cDNA library and increased levels of mRNA for phosphoenolpyruvate carboxylase and pyruvate orthophosphate dikinase. *Photosyn Res* 17: 159-171 (1988).
62. Schuster AM, Davies E: Ribonucleic acid and protein metabolism in pea epicotyls. II Response to wounding in aged tissue. *Plant Physiol* 73: 817-821 (1983).
63. Scott NS, Munns R, Barlow EWR: Polyribosome content in young and aged wheat leaves subjected to drought. *J Expt Bot* 30: 905-911 (1979).
64. Sharp RE, Davies WJ: Solute regulation and growth by roots and shoots of water-stressed maize plants. *Planta* 147: 43-49 (1979).
65. Stewart CR, Voetberg G, Rayapati PJ: The effects of benzyladenine, cycloheximide, and cordycepin on wilting-induced abscisic acid and proline accumulations and abscisic acid- and salt-induced proline accumulation in barley leaves. *Plant Physiol* 82: 703-707 (1986).
66. Takio K, Towatari T, Katunuma N, Teller DC, Titani K: Homology of amino acid sequences of rat liver cathepsins B and H with that of papain. *Proc Natl Acad Sci USA* 80: 3666-3670 (1983).
67. Turner NC: Adaptation to water deficits: A changing perspective. *Aust J Plant Physiol* 13: 175-190 (1986).
68. Williams JG, North MJ, Mahbubani H: A developmentally regulated cysteine proteinase in *Distyostelium discoideum*. *EMBO J* 4: 999-1006 (1985).
69. Williamson JD, Quatrana RS: ABA-regulation of two classes of embryo-specific sequences in mature wheat embryos. *Plant Physiol* 86: 208-215 (1988).
70. Wyn Jones RG, Storey R: Betaines. In: *Physiology and Biochemistry of Drought Resistance in Plants*, pp. 171-204. Academic Press, Australia (1981).

## A tandem of $\alpha$ -tubulin genes preferentially expressed in radicular tissues from *Zea mays*

Lluís Montoliu, Joan Rigau and Pere Puigdomènech

Departamento de Genética Molecular, CID-CSIC, C. Jordi Girona Salgado, 18, 08034 Barcelona, Spain

Received 21 April 1989; accepted in revised form 7 August 1989

**Key words:** cDNA, differential screening, genomic cloning, maize, tandem genes,  $\alpha$ -tubulin

### Abstract

The identification of a cDNA (MR19) corresponding to a maize  $\alpha$ -tubulin and homologous genomic clones (MG19/6 and MG19/14) is described. The cDNA has been isolated by differential screening of a cDNA maize root library. We have found two  $\alpha$ -tubulin genes in a tandem arrangement in the genomic clones, separated by approximately 1.5 kbp. One of the genes (gene I) contains an identical nucleotide sequence which corresponds to the cDNA clone. The two deduced proteins from DNA sequences are very similar (only two conservative replacements in 451 amino acids) and they share a high homology as compared with the published  $\alpha$ -tubulin sequences from other systems and in particular with the *Arabidopsis thaliana* and *Chlamydomonas reinhardtii* sequences reported. The structure of both genes is also very similar; it includes two introns, of 1.7 kbp and 0.8 kbp respectively, in each gene and only one intron placed at a homologous position in relation to *Arabidopsis thaliana* genes. By using specific 3' probes it appears that both genes are preferentially expressed in the radicular system of the plant. The  $\alpha$ -tubulin gene family of *Zea mays* seems to be represented by at least 3 or 4 members.

### Introduction

Microtubules are the basic components of the cytoskeleton. They play an essential role in eukaryotic cells in many processes such as cell division, internal transport, motility and morphogenesis. As compared with animal systems plant cells present particular features that both restrict and give additional functions to microtubules. In fact, plant microtubules have specialized roles in mitosis controlling the plane of cell division and consequently the direction of cell elongation. In this sense microtubules are involved in determin-

ing plant cell morphogenesis [16]. A particular set of plant microtubules, the cortical array, has been proposed to take part in the orientation of cellulose fibrils in the wall [27].

The major component of microtubules is tubulin, a heterodimer protein formed by two different subunits, alpha and beta, of approximately 50 kDa each. Both subunits share 40% of homology. In addition,  $\alpha$ - and  $\beta$ -tubulins have been extremely conserved through evolution. The majority of published sequences can be placed below 15% of divergence [26]. Tubulins ( $\alpha$  and  $\beta$ ) are coded in vertebrates by complex multigene

families. In these systems, members of the family are organized in a dispersed way. In man more than 20  $\alpha$ -tubulin genes have been described and approximately 15 genes have been characterized in mouse or rat [7]. Most of these are supposed to be pseudogenes [25].

In addition, several  $\alpha$ -tubulin sequences have been published in lower eukaryotic systems. More than 15  $\alpha$ -tubulin genes have been described in *Trypanosoma* forming clusters, organized into tandem repeats of alternating  $\alpha$  and  $\beta$  gene units 3.7 kbp long [24].

$\alpha$ -tubulin genes have also been characterized from various photosynthetic organisms such as *Chlamydomonas reinhardtii* [41] and *Volvox carterii* [30]. These two species present two unlinked genes in each case. However, in comparison with the abundant information available for genes coding for  $\alpha$ -tubulin in animal systems, very few data exist from the plant kingdom [42]. Tubulins have been purified from plant cells by using their property of *in vitro* assembly into microtubules [32]. Protein data have indicated the existence of multiple isotypes in tissues of the higher plant *Phaseolus vulgaris* [21], and monocotyledon endosperm cells [35]. Only the  $\alpha$ -tubulin gene family from *Arabidopsis thaliana* has been extensively characterized so far in higher plants [28]. In this species four genes coding for this protein have been detected organized in a dispersed form in the genome. A possible cluster of alternating  $\alpha/\beta$  tubulin genes has been proposed from mung bean [38], but no sequence data from cloning studies are yet available.

The existence of different tubulin genes in a given organism may correspond to the need of a differential expression in defined tissues or developmental stages. Data from the literature indicate that in some cases different genes are expressed in distinct tissues or developmental stages. In rat the presence of a specific cerebellum isoform of  $\alpha$ -tubulin has been described [14] while in mouse two specific genes are expressed in testis [49] and an unusually highly divergent testicular isoform of  $\alpha$ -tubulin, not found in brain, has been described [20]. Constitutive and tissue-specific isoforms of  $\alpha$ -tubulin have been observed in *Drosophila*

*melanogaster* [48]. In contrast with the hypothesis of different  $\alpha$ -tubulin genes present in functionally different microtubules, data from two yeast species studied, *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, show only two genes for  $\alpha$ -tubulin in their genome, only one of which seems to be essentially required whereas the other one is dispensable [43, 1].

As has been shown in animal organisms, plant tubulin genes may be interesting markers to study developmental processes occurring in plants. More complete information available for a plant system comes from *Arabidopsis thaliana*, where one of the genes seems to be specifically expressed in flowers as compared with leaves and roots [29]. Recent reports based on protein data indicate the existence of specific patterns of tubulin polypeptides in different organs of carrot [22]. A developmental control of the expression of tubulin proteins during somatic embryogenesis in cultured carrot cells has been described [9].

In the present paper we show that  $\alpha$ -tubulin genes are cloned in *Zea mays* by a family of genes that includes a subgroup of two genes forming a tandem array. Both genes have a preferential expression in the radicular system. The structure of the proteins is compared with other known systems showing a high level of homology. The novel organization of  $\alpha$ -tubulin genes in this plant species is discussed, as well as in different varieties of the maize group.

## Materials and methods

### Plant material

All our studies were done with *Zea mays* L. (inbred line W64A) grown under greenhouse conditions. Plantlets were obtained by germinating dry seeds through imbibition in water at 25 °C in the dark for the indicated period of time.

### General methods

DNA ligation, bacterial transformation (*Escherichia coli* DH5 $\alpha$ F' strain) and cloning procedures

followed published methods [17]. Harvesting and preparation of lambda-phage DNA were carried out according to described protocols [10] as well as preparation of plasmid DNA, restriction enzyme analysis, agarose gel electrophoresis of nucleic acids and blotting [31]. Vectors used in subcloning and sequencing procedures were pUC18 and M13mp18/19 [50].

Enzymes were purchased from Boehringer (Mannheim) unless stated otherwise, labelled nucleotides were from Amersham.

#### *RNA isolation and poly(A)<sup>+</sup> RNA selection*

20–40 g of plant material were frozen in liquid nitrogen and ground to a fine powder in a mortar. For the construction of cDNA libraries the guanidinium isothiocyanate method was used [31]. For preparing single-strand cDNA probes and Northern blot analysis the RNA was extracted according to described methods [11, 45]. Poly(A)<sup>+</sup> RNA was selected on oligo(dT)-cellulose chromatography.

#### *cDNA library*

A  $\lambda$ gt10 cDNA library was prepared from poly(A)<sup>+</sup> RNA of *Zea mays* roots two months old, essentially according to described protocols [15], with modifications.

1) For first-strand cDNA synthesis 6  $\mu$ g of poly(A)<sup>+</sup> RNA was denatured by incubation at 70 °C for 5 min and the reaction carried out in a volume of 55  $\mu$ L containing 50 mM Tris-HCl pH 8.3 (at 42 °C), 10 mM MgCl<sub>2</sub>, 10 mM DTT, 1 mM dATP, 1 mM dCTP, 1 mM dGTP, 1 mM dTTP, 120 mM KCl, 25 u of RNasin (Stehelin), 5.5  $\mu$ g of oligo-d(T)<sup>12–18</sup> and 20 u/ $\mu$ g RNA of Super Reverse Transcriptase (Stehelin). The mix was incubated at 42 °C for 1 hour.

2) For second-strand cDNA synthesis up to 1  $\mu$ g of single-stranded cDNA (*i.e.* 2  $\mu$ g of hybrid) was processed in 200  $\mu$ L of 20 mM Tris-HCl pH 7.5, 5 mM MgCl<sub>2</sub>, 10 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 100 mM KCl 150  $\mu$ M NAD, 200  $\mu$ M dATP, 200  $\mu$ M dCTP, 200  $\mu$ M dGTP, 200  $\mu$ M dTTP, 50  $\mu$ g/ml BSA, 2.7 u RNase-H (Stehelin), 50 u

DNA Polymerase I (Stehelin), and 5 u *E. coli* DNA ligase (Biolabs). Incubations were sequentially 2 h at 15 °C and 1 h at 22 °C.

3) The cDNA was methylated with *Eco* RI methylase (Biolabs), the ends repaired with T4-DNA-Polymerase I (Klenow, Biolabs), and the cDNA ligated to synthetic *Eco* RI linkers (10-mer, Biolabs). The cDNA attached to the linkers was cleaved with *Eco* RI and fractionated by agarose gel electrophoresis according to length. cDNAs between 500 bp and 3.5 kbp were inserted into the unique *Eco* RI site in the  $\lambda$ cl gene of the  $\lambda$ gt10 [23]. The DNA was packaged *in vitro* [31] and plated onto C600 Hfl *E. coli* K 12 cells. 2  $\mu$ g of poly(A)<sup>+</sup> RNA yielded  $1.2 \times 10^6$  independent clones of  $\lambda$ cl recombinants.

#### *Differential screening of the cDNA library*

Single-strand cDNA probes were obtained with poly(A)<sup>+</sup> RNA from 2-month-old roots and leaves of maize by a modification of the method described previously [46]. The reactions were carried out in a volume of 25  $\mu$ L containing 1  $\mu$ g of poly(A)<sup>+</sup> RNA of each organ separately; 1.4  $\mu$ g of oligo d(T)<sup>12–18</sup> (both prewarmed 3 min at 65 °C and cooled on ice); 50 mM Tris-HCl pH 8.3 (42 °C), 10 mM MgCl<sub>2</sub>, 10 mM DTT, 0.5 mM dATP, 0.5 mM dCTP, 0.5 mM dTTP, 0.03 mM dGTP, 50  $\mu$ Ci <sup>32</sup>P-dGTP (3000 Ci/mmol, Amersham), 30 u RNasin (Genofit) and 10 u of Reverse Transcriptase (Genofit). The mixes were incubated at 42 °C for 45 min and unincorporated <sup>32</sup>P-dGTP separated by chromatography on a 0.7 cm  $\times$  10 cm Sephadex G-50 column. Three plaque filter replicas of 5 plates of the cDNA library (3000 pfu/plate) were prepared. Replicas were obtained by transferring each one during 2, 4 and 8 min onto nitrocellulose paper. The first and third filters were hybridized with single-strand cDNA probes from roots, and the second one with corresponding probes from leaves at a specific activity of ca.  $1.5 \times 10^7$  cpm/ $\mu$ g. Hybridization was performed at T<sub>m</sub>-25 with  $0.6 - 1 \times 10^6$  cpm/ml solution for 24–36 hours, in the presence of 50% formamide, and final washes

(30 min each) were done in  $0.2 \times \text{SSC}$ , 0.1% SDS at 65 °C.

#### *Northern blot analysis*

10  $\mu\text{g}$  of total RNA and 0.5  $\mu\text{g}$  of poly(A)<sup>+</sup> RNA were fractioned in 1.5% agarose-formaldehyde gels [3], transferred to either nitrocellulose (BA85, Schleicher & Schuell, 0.45  $\mu\text{m}$ ) or nylon membranes (Hybond-N, Amersham) and hybridized according to protocols suggested by the suppliers. Probes were labelled by nick-translation [39] or random-primed (Boehringer-Mannheim kit) reactions [12] to a specific activity up to  $10^9$  cpm/ $\mu\text{g}$ .

#### *Southern blot analysis*

Isolation of DNA was performed following procedures described previously [4] with variations. 20 g of plant material were frozen in liquid nitrogen and reduced to powder in a mortar. 4 ml/g of extraction buffer (0.35 M NaCl, 10 mM Tris-HCl pH 8, 0.1 M EDTA, 1% Sarcosyl, 7 M urea and 5% phenol) was added. The mixture was incubated for 2 h at 37 °C under gentle shaking. After phenol/chloroform extraction, DNA was ethanol-precipitated and purified through a CsCl gradient.

10  $\mu\text{g}$  of genomic DNA digested with restriction enzymes were fractioned in 0.8% agarose gels and blotted onto nylon membranes (ZetaProbe, BioRad). Hybridizations were done in:  $1.5 \times \text{SSPE}$  ( $1 \times \text{SSPE}$  is 180 mM NaCl, 10 mM  $\text{NaH}_2\text{PO}_4$  pH 7.7, 1 mM EDTA), 1% SDS, 0.5% powered skimmed milk (Molico, Nestlé) and 0.5 mg/ml of denatured salmon sperm DNA at 65 °C for 16–20 h. Washes were performed also at 65 °C up to  $0.1 \times \text{SSC}$ , 1% SDS.

#### *DNA sequence analysis*

Reactions were carried out using the M13-dideoxy nucleotide method according to de-

scribed protocols [40] with modifications in order to use <sup>35</sup>S-labeled nucleotides [2]. Klenow DNA polymerase (New England Biolabs) and the T7-Sequencing Kit (Pharmacia) were used. The products obtained were separated in 6% acrylamide 8 M urea 0.2–0.45 mm wedge gels. Sequence alignment and analysis was done by using software from the CITI2 database (Paris) and Micro-Genie-(Beckman) [37].

#### *Screening of the genomic library*

The screening of a  $\lambda\text{Ch35}$  genomic library of maize [13] was carried out with cDNA clones labelled by random priming (Boehringer-Mannheim kit). Hybridization conditions of nitrocellulose replicas were:  $10^6$  cpm/ml,  $5 \times \text{SSC}$ ,  $5 \times \text{Denhardt's solution}$ , 50% formamide, 5 mM EDTA, 25 mM sodium phosphate, 0.1% SDS and 250  $\mu\text{g/ml}$  of denatured salmon sperm DNA, at 42 °C for 24 h. The final two washes were done in  $0.2 \times \text{SSC}$ , 0.1% SDS at 65 °C for 15 min each.

### **Results**

#### *cDNA cloning and sequencing*

A group of cDNA clones corresponding to maize  $\alpha$ -tubulin was found in the course of a differential screening of a maize root cDNA library using labelled single-stranded cDNA probes synthesized from root and leaf poly(A)<sup>+</sup> RNA. The library was constructed in  $\lambda\text{gt10}$  and 15 000 recombinant clones were screened. From those giving a preferential signal with root RNA, a number were selected to confirm their predominant expression in the radicular tissues by means of a northern analysis. One of these clones (MR19), giving a 7-fold increase in root DNA as compared to leaf (see Fig. 1A), was one of those chosen for further analysis. The estimated size of the mRNA was approximately 1750 nucleotides.

When northern analysis was carried out with RNA extracted from different parts of the maize plant the preferential expression of MR19 in the

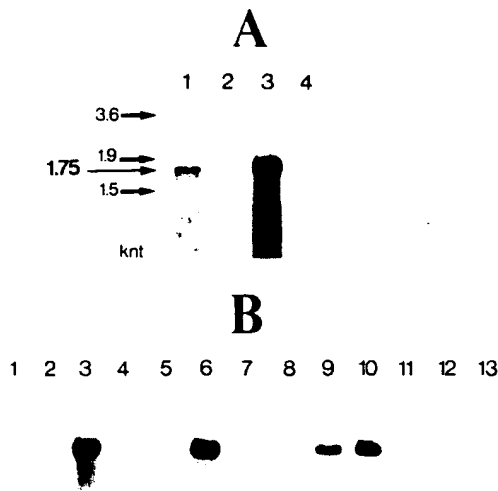


Fig. 1. A. Northern blot analysis of maize adult root and leaf RNA. Total RNAs (10  $\mu$ g) isolated from adult (60 days after germination) roots (1) and leaves (2) and corresponding poly (A)<sup>+</sup> RNAs (0.5  $\mu$ g) fraction from roots (3) and leaves (4) were probed with the insert of MR19 clone (cDNA). The mobilities of 25S, 18S and 16S ribosomal RNAs are indicated. The estimated length of the transcript (in kilonucleotides) is shown. B. Northern blot analysis of MR19  $\alpha$ -tubulin in maize in different organs and several stages of development. The following RNAs were probed with the insert of MR19 (cDNA) clone. Polysomal RNAs, membrane-bound (2  $\mu$ g) purified from endosperm 20 days after pollination (1); total RNAs (10  $\mu$ g) extracted from embryo 20 days after pollination (2); young roots two (3), four (4) and seven (5) days old; root tip from seven-day-old young roots (6); adult root (7); coleoptiles two (8), four (9) and seven (10) days old; upper part of coleoptile (including node) (11) and seven-day-old shoots (12) and adult leaves (13).

radicular tissues was confirmed. The northern analysis of RNA extracted from endosperm and embryo 20 days after pollination, adult (two months old) roots and leaves, and different parts of different ages of the young roots and coleoptiles is presented in Fig. 1B. It appears that, while RNA hybridizing with this cDNA is poorly represented in leaves and shoots, the highest amount of mRNA is found in the meristematic part of the young roots, but it is also found in adult roots. In general, it appears that the mRNA is especially abundant in zones of the young plant rich in meristematic regions, such as young roots and coleoptiles, with a special increase in the radicular

tissues. Although there is a good hybridization signal in coleoptiles, a clear decrease of hybridization in adult leaves is observed as compared with adult roots. Expression is also detected at a lower level in embryo and endosperm.

Sequencing of the insert was carried out with the dideoxy method. The nucleotide sequence and the protein sequence deduced from it are shown in Fig. 2 (between asterisks at positions 6530 and 8210, excluding introns), together with the MG19/6 and MG19/14 genomic clones (described below) in an overlapping form. Searching in the EMBL databank revealed a strong homology with *Chlamydomonas reinhardtii*  $\alpha$ -tubulin [41], and subsequently with all the other published sequences coding for  $\alpha$ -tubulins. As could be expected from the size of the cDNA insert (1483 bp), as compared with the estimated mRNA length, the protein sequence deduced lacks the initial part of the N-terminus of maize  $\alpha$ -tubulin. The non-translated 3' end of the mRNA is 280 nucleotides long and has two possible polyadenylation signals at 27 and 104 nucleotides before the beginning of the poly(A) tail (at position 8210) as shown in Fig. 2. Both sequences (AATAAT and AGTAAA) depart from a single nucleotide in the consensus sequence (AATAAAA).

#### Genomic cloning and sequencing

The cDNA insert of the MR19 clone was used to screen a genomic library constructed in  $\lambda$ Ch35 by partial digestion with Mbo I of endosperm DNA from the maize inbred line W64A [13]. Eleven clones were found positive, and two having a strong hybridization signal (MG19/6 and MG19/14) were used for further analysis. Clones MG19/6 and MG19/14 were found to overlap, MG19/14 being longer at its 3' end and therefore containing the whole cDNA sequence (see Fig. 3). The sequence of the overlapping fragments showed that they corresponded to identical genomic sequences.

From the restriction analysis and partial sequence (8.5 kbp) of the insert of the genomic





10

•

insert and the location of the coding sequences and introns are shown in Fig. 3. The 3' region (gene I) contains a nucleotide sequence identical to the cDNA, except for the presence of three insertions flanked by the consensus sequences of

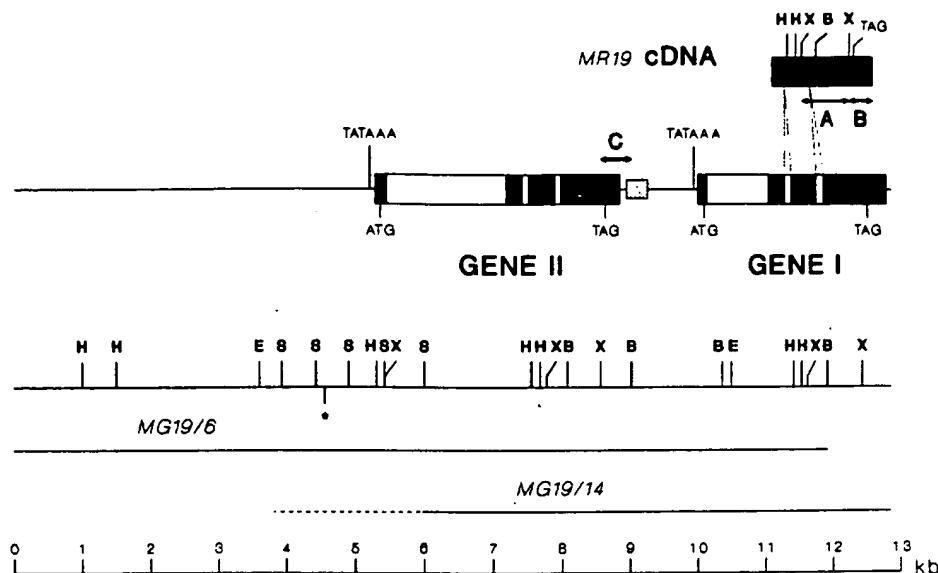


Fig. 3. Restriction maps of genomic clones MG19/6 and MG19/14 showing two  $\alpha$ -tubulin units, genes I and II. In both genes white boxes represent introns and black ones the corresponding exon coding segments. The putative TATA boxes, ATG and stop signals are indicated. The map of the MR19 cDNA clone is also shown. The dotted box located between genes represent a zone rich in repetitive sequences (see Fig. 4). The sites presented here are marked as follows: H (*Hind* III), E (*Eco* RI), S (*Sac* I), X (*Xho* I) and B (*Bam* HI). Arrowed lines marked with A, B and C show probes used in experiments described in Fig. 5 and 6. The nucleotide sequence presented in Fig. 2 lies between a *Sal* I site (marked with an asterisk) and the 3' end of the MG19/14 genomic clone.

introns. The central region of the genomic clone, hybridizing with the MR19 probe, resulted in another gene (gene II) coding for an  $\alpha$ -tubulin protein of an identical protein sequence as compared with the other one, except for two conservative replacements (Glu/Asp, Asp/Glu) in the hypervariable acidic C-terminus of the protein (see Fig. 7).

The nucleotide sequence of the two genes is very similar but clearly not identical. They show 95% homology when only the translated sequences are considered. Among 1353 translated nucleotides in both genes we have found 72 nucleotide replacements while only 2 (as mentioned before) result in amino acid changes.

The sequence of the unit of tandem genes, resulting from MG19/6, MG19/14 and MR19 clones, is presented in Fig. 2. Genes I and II are interrupted by three introns. The first and third introns are located between amino acid positions 31/32 and 233/234, respectively, whereas the second one is found between the first and second

base of triplet coding for amino acid 110 of the protein, the same position of the second intron of *A. thaliana*  $\alpha$ -tubulin genes [42]. In particular, although their position is identical in the two genes, the length and the sequence of the introns are quite different. The first intron is 859 bp long in gene I and 1724 bp in gene II, the second one is 93 bp long in both cases, and the third one is 105 bp long in gene I and 92 bp long in gene II. The six introns present consensus 5' (GT-) and 3' (-AG) recognition signals as reported in the literature [18].

Both genes have consensus TATA boxes at approximately the same distance from the initial ATG, which is -149 and -152 for gene I and II respectively. The nucleotide positions surrounding the putative TATA boxes have been conserved showing a high degree of homology. When the 5' end of the coding region is compared with the  $\alpha$ 3-tubulin *A. Thaliana* sequence [28], good homology is found between the transcriptional starting point for gene  $\alpha$ 3 of *Arabidopsis* and the

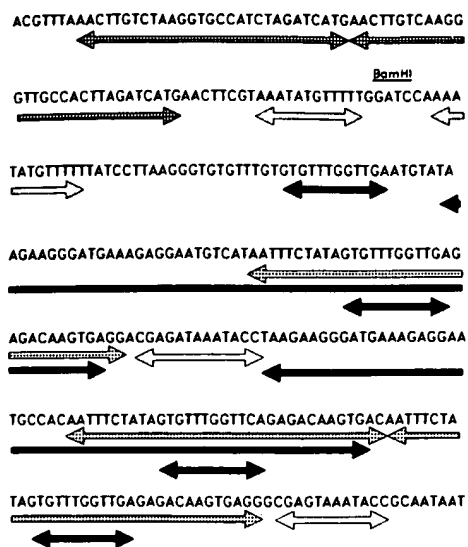


Fig. 4. Repetitive sequences found in the intergenic region of the MG19/6 genomic clone. Equal patterns of duplication are shown with homologous filled arrows. The Bam HI site is presented to locate the region in the genomic map (see Fig. 3).

maize genes. This point is situated at 34 bp (gene II), and 35 bp (gene I), downstream from the putative TATA boxes. Both genes have TAG as stop codon and present possible polyadenyla-

tion signals (shown in Fig. 2) placed, as mentioned before, in gene I and a single one at 120 nucleotides from the stop codon in gene II.

The intergenic region is 1452 bp long, from the stop codon of gene II to the ATG of gene I. In this zone homologies between the two genes are found in the transcribed 3' and 5' ends and within an approximately 200 bp region surrounding the TATA box. An especially interesting sequence is found in a zone (marked with a dotted box in Fig. 3 and denoted with dots in Fig. 2) that seems to separate the two genic regions. A number of repetitions not observed in other parts of the sequenced fragment is found. Repeats from 12 up to 60 bp long can be observed in a variable number up to 4 times. This sequence is shown in Fig. 4 with the indicated repetitions.

#### Gene structure and differential expression

The proteins encoded by gene I and II slightly differ in their C-terminus. the 3' untranslated region has also marked differences between the two genes. As has been done in other cases using

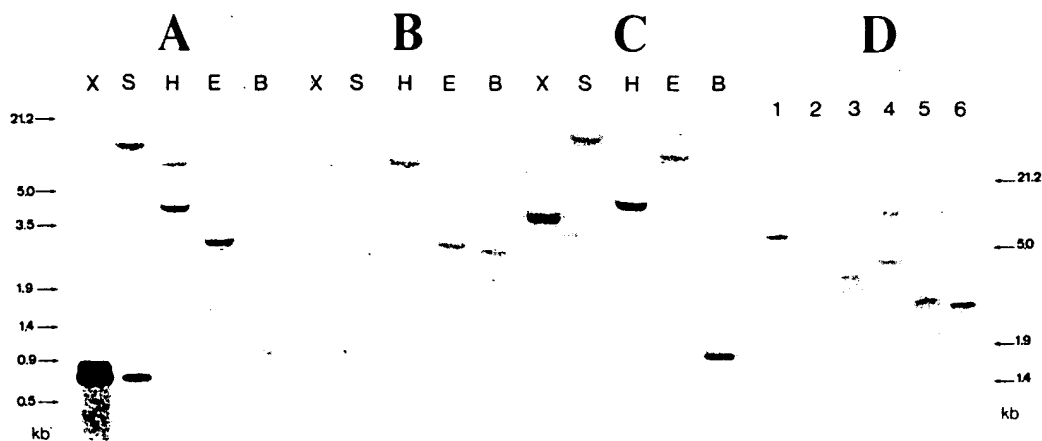


Fig. 5. Southern analysis of genomic  $\alpha$ -tubulin sequences homologous to the MR19 cDNA clone of *Zea mays*. Southern blots were prepared from *Zea mays* genomic DNA digested with *Xho* I (X), *Sac* I (S), *Hind* III (H), *Eco* RI (E) and *Bam* HI (B) and hybridized with A, B and C probes respectively (see Fig. 3). Probe A corresponds to the most conserved part of  $\alpha$ -tubulin sequences. Probes B and C correspond to the hypervariable and specific carboxyl-terminus protein sequences and the 3' ends of genes I and II. The same filter was reused in order to allow a better comparison of the observed signals. D. Southern blot analysis of  $\alpha$ -tubulin genes in several related cereals. DNAs were extracted from *Sorghum bicolor* (1), *Zea diploperennis* (teosinte) (2), an F1 from *Zea diploperennis* x *Zea mays* (Palomero Toluqueño) (3), *Zea mays* inbred line A188 (4), *Zea mays* inbred line W64A (5) and *Zea mays* Black Mexican Sweet (6). All these DNAs were digested with *Eco* RI and hybridized with probe A (see Fig. 3).

3' specific probes, it could be possible to study both the genomic structure as seen in Southern analysis and the expression of the two genes by northern blots. The Southern blot using a probe that contains the nucleotide region coding for the most conserved domain of the protein (probe A in Fig. 3) gives the pattern shown in Fig. 5A.

At this level of stringency two types of bands of different intensity may be observed (see, for instance, the *Hind* III lane). Using 3' specific probes (probes B and C in Fig. 3) a pattern giving rise to single Southern bands is obtained (Fig. 5B and C). By comparing the three blots it is possible to interpret the pattern of intense bands obtained with the probe of the invariable zone and, with the help of the two other probes, to attribute them to either gene I or gene II (see Fig. 3).

The bands observed in the Southern blots using 3' probes correspond exactly to the genomic restriction fragments of the cloned segments as presented in Fig. 3, except where sites outside the clones are involved. The *Xho* I lane shows three bands when probe A is used. The lowest and most intense one corresponds to both gene I and gene II (see restriction map in Fig. 3), giving a higher intensity of hybridization. The bands having a lower mobility and intensity are supposed to correspond to other members of the family of  $\alpha$ -tubulin genes. *Sac* I lane gives only two bands using probe A. The larger one would match perfectly to both corresponding genomic fragments of genes I and II, as can be seen when probes B and C are used, whereas the smaller one will be the sequence homologous to other gene(s). The *Eco* RI lane shows the simplest pattern since using probe A only two bands appear. Each one can be directly related to gene I or gene II when specific probes are used. The *Hind* III lane gives two bands of higher intensity together with two or three smaller and fainter bands. The first and second bands account for the corresponding genes (I or II, observing the signal when specific probes are used). The rest of the bands correspond to the other less homologous members of the family. A similar behaviour can be observed in *Bam* HI lanes.

The genomic structure of the corresponding

genes of  $\alpha$ -tubulin in different varieties of maize and related species has also been studied. In Fig. 5D the Southern blot of DNA from *Sorghum bicolor*, *Zea diploperennis* (teosinte), and F1 generation resulting from *Zea diploperennis*  $\times$  *Zea mays* (Palomero toluqueño) and three inbred lines of *Zea mays* (A188, W64A and Black Mexican Sweet) is shown. Although polymorphisms are observed at the genomic level, the number of genes homologous to those here studied appear to be similar in all these plants.

The same probes used for the Southern analysis were used to measure the expression of the two genes in different parts of the plant. The pattern of distribution of the mRNA using the probe for the conserved region (Fig. 6A) or the 3' probe of gene I (Fig. 6B) is quite similar. When the 3' probe of gene II is used (Fig. 6C) some dif-

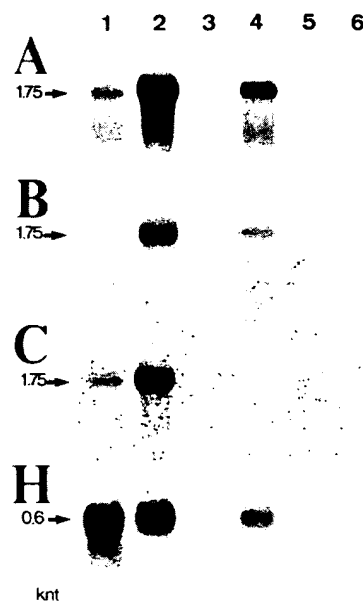


Fig. 6. Analysis of mRNA accumulation of  $\alpha$ -tubulin genes I and II. Northern blots were prepared with total RNAs extracted from, two-day-old coleoptiles (1) and young roots (2); poly(A)<sup>+</sup> RNA (0.5  $\mu$ g) from adult leaves (3) and roots (4); total RNAs (10  $\mu$ g) from adult leaves (5) and roots (6). The filters were hybridized with probes A, B and C, respectively (see Fig. 3) and with histone H3 maize probe (H). Filters were washed and reprobated for a better comparison of results. In each case the same time of exposition (7 days) was used.

ferences can be observed. In particular, the expression of this gene seems to be only confined to young plantlets with preferential expression in young roots.

Parallel experiments were carried out with maize histone H3 [6] (shown in Fig. 6) and H4 [34] (data not shown), as probes for detecting expression in organs rich in dividing cells. The results indicated that a very similar pattern of mRNA accumulation of these two histone genes is obtained in organs rich in meristematic regions, such as young roots and coleoptiles, while a different pattern is observed when probe A, B or C are used. Autoradiographs of northern blots were scanned in a densitometer to obtain relative quantification of bands in the different lanes. When the signal of histone H3 is taken as a reference, the expression of gene I in radicular tissues is 9 times more intense than in the coleoptile, and there is a 6-fold increase in gene II indicating a preferential expression of gene I (and II) in the radicular system of *Zea mays*.

## Discussion

A tandem repeat of  $\alpha$ -tubulin genes has been cloned in maize as a result of a differential

screening of a root cDNA library. The two genes code for two almost identical proteins with only two conservative (Asp/Glu and Glu/Asp) replacements in the hypervariable C-terminus of the protein. The sequences are more variable at the nucleotide level but keep 95% homology between the two genes.

When the protein sequence of maize  $\alpha$ -tubulin is compared with that of other known  $\alpha$ -tubulin sequences, a high degree of similarity is found. In Fig. 7 the two maize sequences are compared with the two  $\alpha$ -tubulin sequences known from *A. thaliana*,  $\alpha 1$  [29] and  $\alpha 3$  [28], with the unicellular algal sequence  $\alpha 1$ -tubulin from *Chlamydomonas reinhardtii* [41], and with the  $\kappa \alpha 1$ -tubulin from man [8]. The homology between two organisms as diverged as *Homo sapiens* and *Zea mays* yields a remarkable level of homology (83.4%).

Comparison of the protein sequences demonstrate that the maize sequences  $\alpha 1$ - and  $\alpha 2$ -tubulin corresponding to genes I and II, respectively, share an 99.6% homology. In addition, the highest similarity is found between the *Chlamydomonas* protein and  $\alpha 1$ -maize (91.8%) or  $\alpha 2$ -maize (91.6%), a degree of homology exceeding that found with  $\alpha 3$  protein from *Arabidopsis* (91.2% for  $\alpha 1$ -maize and 90.7% for  $\alpha 2$ -maize). Homology

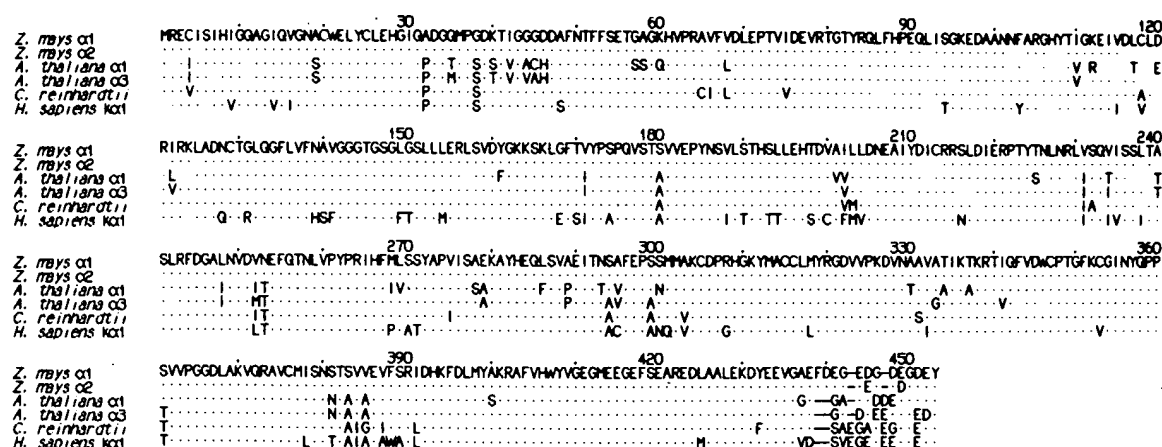


Fig. 7. Comparison of  $\alpha$ -tubulin protein sequences between different organisms. The predicted amino acid sequence of the  $\alpha 1$ -tubulin of *Zea mays* (coded by gene I) is presented at the top line of single-letter code. The predicted amino acid sequences of the  $\alpha 2$ -tubulin of *Zea mays* (coded by gene II) and other  $\alpha$ -tubulin are shown below respecting an optimal alignment. Only those amino acids that change from  $\alpha 1$ -tubulin of *Zea mays* are indicated. A dash in the sequence means a gap in the amino acid sequence introduced in order to maintain a good alignment. Sequences from *Chlamydomonas reinhardtii*  $\alpha 1$ -tubulin, *Arabidopsis thaliana*  $\alpha 1$  and  $\alpha 3$  tubulin, and *Homo sapiens*  $\kappa \alpha 1$ -tubulin are those published in refs. [41], [29], [28] and [8], respectively.

between  $\alpha 1$  and  $\alpha 3$  proteins of *Arabidopsis* is lower (91.2%) than between the two maize tubulins. These data would suggest that the duplication observed in maize is relatively recent and, in any case, more recent than the duplication observed in *Arabidopsis*. On the other hand, we observe by Southern analysis a similar pattern of genomic structure in different varieties of *Zea mays* and related species (Fig. 5D), indicating that the existence of these two related genes may be a general characteristic of the group.

Both deduced proteins  $\alpha 1$  and  $\alpha 2$  from maize present a Tyr as a C-terminal end. The same has been reported in all other plant or algal  $\alpha$ -tubulin sequences published showing that the phenomenon of absence of this ending amino acid of the variable 3' terminus may be restricted to the animal kingdom [36]. There is also a high degree of conservation in the sequence postulated to be involved in GTP binding in the  $\alpha$ -tubulin sequences (amino acids 143–149) [44].

The introns and the flanking regions show a much larger divergence between the two genes than the coding sequence. The similarity between homologous introns of both gene I and II are 88.2% in the second intron and 61.9% in the third one. A much lower similarity in the first intron is observed. It presents homology only at defined stretches. This intron seems to have evolved incorporating different sequences in gene II which are not present at all in the corresponding region of gene I. This first intron is unusually long (859 bp for gene I, 1724 bp for gene II) for plant introns reported until now [19]. Likewise, the location of introns differs from that observed in other species. Only the second intron is placed at the same amino acid position (110) as those observed in the two  $\alpha$ -tubulin genes of *Arabidopsis thaliana* [28, 29], while the two other ones are different in position and length. Although the position is the same in the second intron their lengths are quite different in both plants (512 bp for the  $\alpha 1$  gene, and 76 bp for the  $\alpha 3$  gene of *A. thaliana*; 93 bp in both genes of *Z. mays*). Low homologies are observed when the second introns of both organisms are compared, but short segments (up to 30 nt) of the sequence appear to be extremely

similar. The finding of the second intron conserved at the same place in both plants, *A. thaliana* and *Z. mays*, may be a result of the obvious evolutionary relationships between both organisms, while the homology observed when the coding region is compared puts the described maize  $\alpha$ -tubulin genes nearer to the corresponding ones from *C. reinhardtii* (91.7% on the average) than to the other ones from *A. thaliana* (88.2% on the average). These results may reflect the existence of two different rates of evolution, one related with intron positions and the other one corresponding to the coding sequence itself.

In the flanking regions homology is found between the two genes in the region surrounding the TATA boxes up to -250 from the ATG. When more upstream regions are compared defined homologous segments can be observed in both genes separated by divergent sequences. An interesting sequence is found in the 5' end of the intergenic region. This is the only part of the 8.5 kbp sequence where a high degree of repetitions is found. In fact, a complex structure of duplications may be observed at this point (see Fig. 4). The functional meaning of these duplications either as a part of the promoter or as a hinge region in the gene duplication will be object of future experiments.

The cDNA clone presented has been identified after a differential screening of a root cDNA library searching for clones corresponding to a mRNA preferentially expressed in the radicular tissues. The northern analysis presented indicate that the  $\alpha$ -tubulin genes described here are expressed in the meristematic parts of the plant, for example in the coleoptile or in the root tip, and therefore their expression may be correlated with cell division. In order to investigate this correlation we decided to use maize histone probes [6, 34] as a reference for division activity, since H3 and H4 genes are particularly expressed where DNA replication is present. In addition, it has been proposed that histone and tubulin genes have a cell-cycle regulation, giving the major peaks of accumulation simultaneously at the late G2 phase, as has been demonstrated in *Physarum* [5]. However, when the mRNA level of  $\alpha$ -tubulin

is compared with that of histone genes a relatively high expression is observed in the radicular tissues, indicating that what is observed is not only the result of a concentration of dividing cells (as could result from the differential screening approach) but also an organ-specific control superimposed on these one. The two genes show similar patterns of expression with gene II mRNA accumulation being restricted to the meristematic parts of the young plantlet.

Tubulin genes expressed in defined tissues have been observed in animals and data are available from human, mouse, rat, chicken and other vertebrate and invertebrate organisms [47]. In vertebrates, the variable acidic C-terminus amino acid positions of the protein would act as a specific domain related to the particular expression of the gene. No evidence in that direction has been detected in plant tubulin sequences, probably due to the lower number of sequences and situations reported. In plants little is known about this question. Most data come from *A. thaliana* where a gene, the  $\alpha 1$ -tubulin [29], is differentially expressed in flowers and through different phases of flowering, whereas the  $\alpha 3$ -tubulin [28] is expressed constitutively in all organs examined. Also a  $\beta$ -tubulin (the  $\beta 1$  gene from *A. thaliana*) has been described presenting a preferential accumulation in the root system of the plant as we have shown for the  $\alpha 1$  and  $\alpha 2$  maize tubulin genes. In addition, a possible organization in a tandem array has been proposed for other  $\beta$ -tubulin genes of *A. thaliana*, but no sequence data are available up to now [33].

In summary, a group of  $\alpha$ -tubulin genes from maize have been cloned whose members (genes I and II) are highly homologous. These two genes are organized in a tandem arrangement. This kind of organization for the  $\alpha$ -tubulin genes has not been reported before in the genome of higher eukaryotic organisms, except for some clusters of  $\alpha$ - and  $\beta$ -tubulin genes present in sea urchin [7].

When specific probes for gene I (probe B) and gene II (probe C) are used, a single-gene pattern can be deduced since in all five lanes (in both genes) a single band is observed (Fig. 5). If the bands observed when using those specific probes

would correspond to more than one gene it would imply an extremely high conservation of sequences, even of those located downstream and upstream of the coding regions of the  $\alpha$ -tubulin genes. That would be unlikely in maize, specially when polymorphisms are present in most of the restriction sites used, located within the coding region of these two highly conserved genes. Southern analysis also shows that other  $\alpha$ -tubulin sequences, having a lower degree of homology, exist in the genome of the plant. Our results and analysis of different genomic clones obtained (data not shown) would indicate a small number of genes (probably three or four) coding for the whole  $\alpha$ -tubulin family of *Zea mays*.

The expression of both  $\alpha$ -tubulin genes could be divided into two aspects which may interact at different levels of regulation of gene expression. First, it is clear that a divisional control exists, since both histone (H3/H4) and  $\alpha$ -tubulin genes accumulate in the same meristematic organs such as root tips and coleoptiles. In addition, a second level of control appears indicating a preferential expression in radicular tissues when the signal of histone H3 probe is taken as a reference.

It would be very interesting to focus future experiments on analyzing the meaning of this kind of organ-specific expression, in order to locate the  $\alpha 1$ - and  $\alpha 2$ -tubulin proteins in particular sets of cells. In this way, genes I and II and their protein products may be useful markers of development in the root system of maize.

### Acknowledgements

The authors are grateful to Dr J Avila (CBM-CSIC, Madrid) for revising the manuscript, to Dr C Gigot (IBMP, Strasbourg, France) for providing H3 and H4 probes, to Ms R Raz, Dr M Torrent, Dr V Stiefel and Mr J Gomez for generous help in different aspects of this work, and specially to Mr Patri Gomez for help in sequencing techniques. This work was supported by grants from CSIC (No 500), EECC (B.A.P.) and CIRIT (L.M., 1987). L.M. is a recipient of a fellowship from PFPI.



## References

- Adachi Y, Toda T, Niwa O, Yanagida M. Differential expressions of essential and nonessential  $\alpha$ -tubulin genes in *Schizosaccaromyces pombe*. *Mol Cell Biol* 6: 2168–2178 (1986).
- Biggin MD, Gibson TJ, Hong GF. Buffer gradient gels and  $^{35}\text{S}$  label as an aid to rapid DNA sequence determination. *Proc Natl Acad Sci USA* 80: 3963–3965 (1983).
- Boedtker H. Conformation independent molecular weight determinations of RNA by gel electrophoresis. *Biochim Biophys Acta* 240: 448–453 (1971).
- Burr B, Burr FA. Controlling element events at the shrunken locus in maize. *Genetics* 98: 143–156 (1981).
- Carrino JJ, Laffter TG. Transcription of  $\alpha$ -tubulin and histone H4 begins at the same point in the *Physarum* cell cycle. *J Cell Biol* 102: 1666–1670 (1986).
- Chaubet N, Philipps G, Chaboute ME, Ehling M, Gigot C. Nucleotide sequence of two corn histone H3 genes. Genomic organization of the corn histone H3 and H4 genes. *Plant Mol Biol* 6: 253–263 (1986).
- Cleveland DW, Sullivan KF. Molecular biology and genetics of tubulin. *Ann Rev Biochem* 54: 331–365 (1985).
- Cowan NJ, Dobner PR, Fuchs EV, Cleveland DW. Expression of human alpha-tubulin genes: interspecies conservation of 3' untranslated regions. *Mol Cell Biol* 3: 1738–1745 (1983).
- Cyr RJ, Bustos MM, Guiltinan MJ, Fosket DE. Developmental modulation of tubulin protein and mRNA levels during somatic embryogenesis in cultured carrot cells. *Planta* 171: 365–376 (1987).
- Davis RW, Bolstein D, Roth RJ. A Manual for Genetic Engineering. Advanced Bacterial Genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1980).
- Dean C, Van den Elzen P, Tanaki S, Dunsmuir P, Bedbrook J. Differential expression of the eight genes of the petunia rubulose biphosphate carboxylase small subunit multi-gene family. *EMBO J* 4: 3055–3061 (1985).
- Feinberg AP, Vogelstein B. A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem* 132: 6–13 (1983).
- Gallardo D, Reina M, Rigau J, Boronat A, Palau J. Genomic organization of the 28 kDa glutelin-2 gene from maize. *Plant Sci* 54: 211–218 (1988).
- Ginzburg I, Teichman A, Griffin WST, Littauer UZ. Differential expression of  $\alpha$ -tubulin mRNA in rat cerebellum as revealed by *in situ* hybridization. *FEBS Lett* 194: 161–164 (1986).
- Gubler U, Hoffman BJ. A simple and very efficient method for generating cDNA libraries. *Gene* 25: 263–269 (1983).
- Gunning BES, Hardham AR. Microtubules. *Ann Rev Plant Physiol* 33: 651–698 (1982).
- Hanahan D. Techniques for transformation of *E. coli*. In: Glover DM (ed) DNA Cloning – A Practical Approach, pp. 109–137. IRL press, Oxford (1985).
- Hanley BA, Schuler MA. Plant intron sequences: evidence for distinct groups of introns. *Nucleic Acids Res* 16: 7159–7175 (1988).
- Hawkins JD. A survey on intron and exon lengths. *Nucleic Acids Res* 16: 9893–9908 (1988).
- Hecht NB, Distel RJ, Yelick PC, Tanhauser SM, Driscoll CE, Goldberg E, Tung KSK. Localization of a highly divergent mammalian testicular  $\alpha$ -tubulin that is not detectable in brain. *Mol Cell Biol* 8: 996–1000 (1988).
- Hussey PJ, Gull K. Multiple isotypes of  $\alpha$ - and  $\beta$ -tubulin in the plant *Phaseolus vulgaris*. *FEBS Lett* 181: 113–118 (1985).
- Hussey PJ, Lloyd CW, Gull K. Differential and developmental expression of  $\beta$ -tubulins in a higher plant. *J Biol Chem* 263: 5474–5479 (1988).
- Huynh TU, Young RA, Davis RW. Constructing and screening cDNA libraries in lambda-gt10 and lambda-gt11. In: Glover DM (ed) DNA cloning – A Practical Approach, pp. 49–78. IRL Press Oxford (1985).
- Kimmel BE, Samson S, Wu J, Hirschberg R, Yarbrough LR. Tubulin genes of the African trypanosome *Trypanosoma brucei rhodesiense*: nucleotide sequence of a 3.7 kbp fragment containing genes for alpha and beta tubulins. *Gene* 35: 237–248 (1985).
- Lemischka I, Sharp PA. The sequences of an expressed rat  $\alpha$ -tubulin gene and a pseudogene with an inserted repetitive element. *Nature* 300: 330–335 (1982).
- Little M, Seehaus T. Comparative analysis of tubulin sequences. *Comp Biochem Physiol* 90: 655–670 (1988).
- Lloyd CW. The Plant Cytoskeleton: The impact of fluorescence microscopy. *Ann Rev Plant Physiol* 38: 119–139 (1987).
- Ludwig SR, Oppenheimer DG, Silflow CD, Snustad DP. Characterization of the  $\alpha$ -tubulin gene family of *Arabidopsis thaliana*. *Proc Natl Acad Sci USA* 84: 5833–5837 (1987).
- Ludwig SR, Oppenheimer DG, Silflow CD, Snustad DP. The  $\alpha$ 1-tubulin gene of *Arabidopsis thaliana*: primary structure and preferential expression in flowers. *Plant Mol Biol* 10: 311–321 (1988).
- Mages W, Salbaum JM, Harper JF, Schmilt R. Organization and structure of *Volvox*  $\alpha$ -tubulin genes. *Mol Gen Genet* 213: 449–458 (1988).
- Maniatis T, Fritsch EF, Sambrook J. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1982).
- Morejohn LC, Fosket DE. Higher plant tubulin identified by self-assembly into microtubules *in vitro*. *Nature* 297: 426–428 (1982).
- Oppenheimer DG, Haas N, Silflow CD, Snustad DP. The  $\beta$ -tubulin gene family of *Arabidopsis thaliana*: preferential accumulation of the  $\beta$ 1 transcript in roots. *Gene* 63: 87–102 (1988).
- Philipps G, Chabet N, Chaboute ME, Ehling M, Gigot C. Genomic organization and nucleotide sequences of two corn histone H4 genes. *Gene* 42: 225–229 (1986).

35. Picquot P, Lambert AM. Tubulin from monocotyledon endosperm cells: Particular biochemical and immunological properties. *J Plant Physiol* 132: 561-568 (1988).
36. Pratt LF, Okamura S, Cleveland DW. A divergent testis-specific  $\alpha$ -tubulin isotype that does not contain a dosed C-terminal tyrosine. *Mol Cell Biol* 7: 552-555 (1987).
37. Queen C, Korn LJ. A comprehensive sequence analysis program for the IBM personal computer. *Nucleic Acids Res* 12: 581-599 (1984).
38. Raha D, Sen K, Biswas BB. cDNA cloning of  $\beta$ -tubulin gene and organisation of tubulin genes in *Vigna radiata* (mung bean) genome. *Plant Mol Biol* 9: 565-571 (1987).
39. Rigby PWJ, Diekmann M, Rhodes C, Berg P. Labelling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. *J Mol Biol* 113: 237-251 (1977).
40. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 74: 5463-5467 (1977).
41. Silfow CD, Chisholm RL, Conner TW, Ranum LPW. The two alpha-tubulin genes of *Clamydomonas renhardtii* code for slightly different proteins. *Mol Cell Biol* 5: 2389-2398 (1985).
42. Silfow CD, Oppenheimer DG, Kopczak SD, Ploense SE, Ludwig SR, Haas N, Snustad PD. Plant tubulin genes: structure and differential expression during development. *Dev Genet* 8: 435-460 (1987).
43. Schatz PJ, Solomon F, Botstein D. Genetically essential and nonessential  $\alpha$ -tubulin genes specify functionally interchangeable proteins. *Mol Cell Biol* 6: 3772-3733 (1986).
44. Sternlicht H, Yaffe MB, Farr GW. A model of the nucleotide-binding site in tubulin. *FEBS Lett* 214: 226-235 (1987).
45. Stiefel V, Perez-Grau LI, Albericio F, Giralt E, Ruiz-Avila L, Ludevid MD, Puigdomenech P. Molecular cloning of cDNAs encoding a putative cell wall protein from *Zea mays* and immunological identification of related polypeptides. *Plant Mol Biol* 11: 483-493 (1988).
46. St John TP, Davis RW. Isolation of galactose inducible DNA sequences from *Saccharomyces cerevisiae* by differential plaque filter hybridization. *Cell* 16: 443-452 (1979).
47. Sullivan KF. Structure and utilization of tubulin isotypes. *Ann Rev Cell Biol* 4: 687-716 (1988).
48. Theurkauf WE, Baum H, Bo J, Wensink PC. Tissue-specific and constitutive  $\alpha$ -tubulin genes of *Drosophila melanogaster*. *Proc Natl Acad Sci USA* 83: 8477-8481 (1986).
49. Villasante A, Wang D, Dobner P, Dolph P, Lewis SA, Cowan NJ. Six mouse  $\alpha$ -tubulin mRNAs encode five distinct isotypes: testis-specific expression of two sister genes. *Mol Cell Biol* 6: 2409-2419 (1986).
50. Yanisch-Pervon C, Vieira J, Messing J. Improved M13 phage cloning and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* 33: 103-119 (1985).

W6K

# Absciscic acid and water-stress induce the expression of a novel rice gene

Claim 61

ABA-inducible promoter

John Mundy and Nam-Hai Chua

Laboratory of Plant Molecular Biology, Rockefeller University, 1230 York Avenue, New York, NY 10021, USA

Communicated by D. von Wettstein

We have identified a novel rice gene, called RAB 21, which is induced when plants are subject to water-stress. This gene encodes a basic, glycine-rich protein (mol. wt 16 529) which has a duplicated domain structure. Immunoblots probed with antibodies raised against  $\beta$ -galactosidase/RAB 21 fusion protein detect RAB 21 protein only in cytosolic cell fractions. RAB 21 mRNA and protein accumulate in rice embryos, leaves, roots and callus-derived suspension cells upon treatment with NaCl (200 mM) and/or the plant hormone abscisic acid (10  $\mu$ M ABA). The effects of NaCl and ABA are not cumulative, suggesting that these two inducers share a common response pathway. Induction of RAB 21 mRNA accumulation by ABA is rapid (< 15 min in suspension cells) and does not require protein synthesis, indicating that pre-formed nuclear and/or cytosolic factors mediate the response to this hormone. We have characterized the RAB 21 gene by determining the complete nucleotide sequence of a nearly full-length cDNA and corresponding genomic copy, and by mapping the start site of its major transcript. The proximal promoter region contains various GC-rich repeats.

**Key words:** genomic sequence/cDNA sequence/cycloheximide/osmotic regulation/water deficit

## Introduction

The hormone abscisic acid (ABA) mediates a number of important physiological processes in plants (King, 1976; Jones *et al.*, 1987). Developmental studies have shown that ABA induces the accumulation of specific mRNAs and proteins late during embryogenesis in seeds of diverse species (Galau *et al.*, 1986; Finklestein *et al.*, 1985; Litts *et al.*, 1987). At this time, the level of endogenous ABA increases, the seeds desiccate and the embryos of some species become dormant (King, 1976; Suzuki *et al.*, 1981). In cereal seeds, some of the 'late' developmental mRNAs are long-lived in mature, dry grains, but are rapidly degraded during seed germination. However, exogenously applied ABA causes the precocious accumulation of these mRNAs in immature embryos and their reappearance in germinating seeds (Finklestein *et al.*, 1985; Mundy *et al.*, 1986). Little is known about the functions of these ABA-inducible proteins or about their intracellular localization. Some of them are storage polypeptides (Finklestein *et al.*, 1985; Bray and Beachy, 1985),

while others, such as lectins (Raikel and Wilkins, 1987) or an enzyme inhibitor (Mundy *et al.*, 1986) may be involved in seed protection and/or the maintenance of dormancy.

Physiological studies have shown that endogenous ABA levels increase in plant tissues subjected to water-stress by high osmoticum, NaCl, or drying (Henson, 1984; Jones *et al.*, 1987). Under these conditions, specific mRNAs and proteins accumulate which could affect intra-cellular osmolarity or have other protective functions (Finklestein and Crouch, 1986; Ramagopal, 1987). One such salt-inducible protein whose accumulation is increased by ABA has been isolated from tobacco and shown to be homologous to members of a group of proteinase inhibitors (Singh *et al.*, 1987; Richardson *et al.*, 1987). These results suggest that some of the ABA-inducible mRNAs and proteins which accumulate during seed desiccation are part of a general response by the plant to water deficit. If so, then part of this response involves the synthesis of enzyme inhibitors and lectins which may protect plant tissues from degradation by pathogens during periods of arrested growth and development.

Both the developmental studies on seeds and the physiological studies on water-stress indicate that ABA controls the accumulation of specific mRNAs and proteins. However, it is unclear whether ABA acts at the transcriptional or post-transcriptional level, or both (Jacobsen and Beach, 1986; Mozer, 1980). The mode of action of the hormone via receptors and/or transducing pathways also remains obscure (Hornberg and Weiler, 1984) and it is not known whether transduction of the ABA response signals requires *de novo* protein synthesis. To date, no genomic sequences have been reported for plant genes strongly induced by ABA. Characterization of promoter sequences of such genes will provide a tool with which to dissect the mechanism by which ABA regulates specific gene expression.

We are interested in determining how gene expression is regulated by plant hormones. We have chosen to study the effect of ABA on gene expression in rice because, as we demonstrate here, this hormone plays a central role in seed development and in the response of rice plants to water-stress, two important agronomic traits (Chang *et al.*, 1986; Seshu and Sorrells, 1986). Knowledge of the structure and function of ABA-responsive proteins will aid our understanding of the physiology of seed maturation and of drought tolerance in cereals. As a first step, we have isolated several cDNA clones whose expression is induced by ABA and water-stress. One of these clones, called RAB 21 (for Responsive to ABA), was fully characterized. We present here the sequence of this novel rice gene, and a characterization of the RAB 21 protein product. We show that the induction of the RAB 21 by ABA and water-stress is rapid and independent of *de novo* protein synthesis.

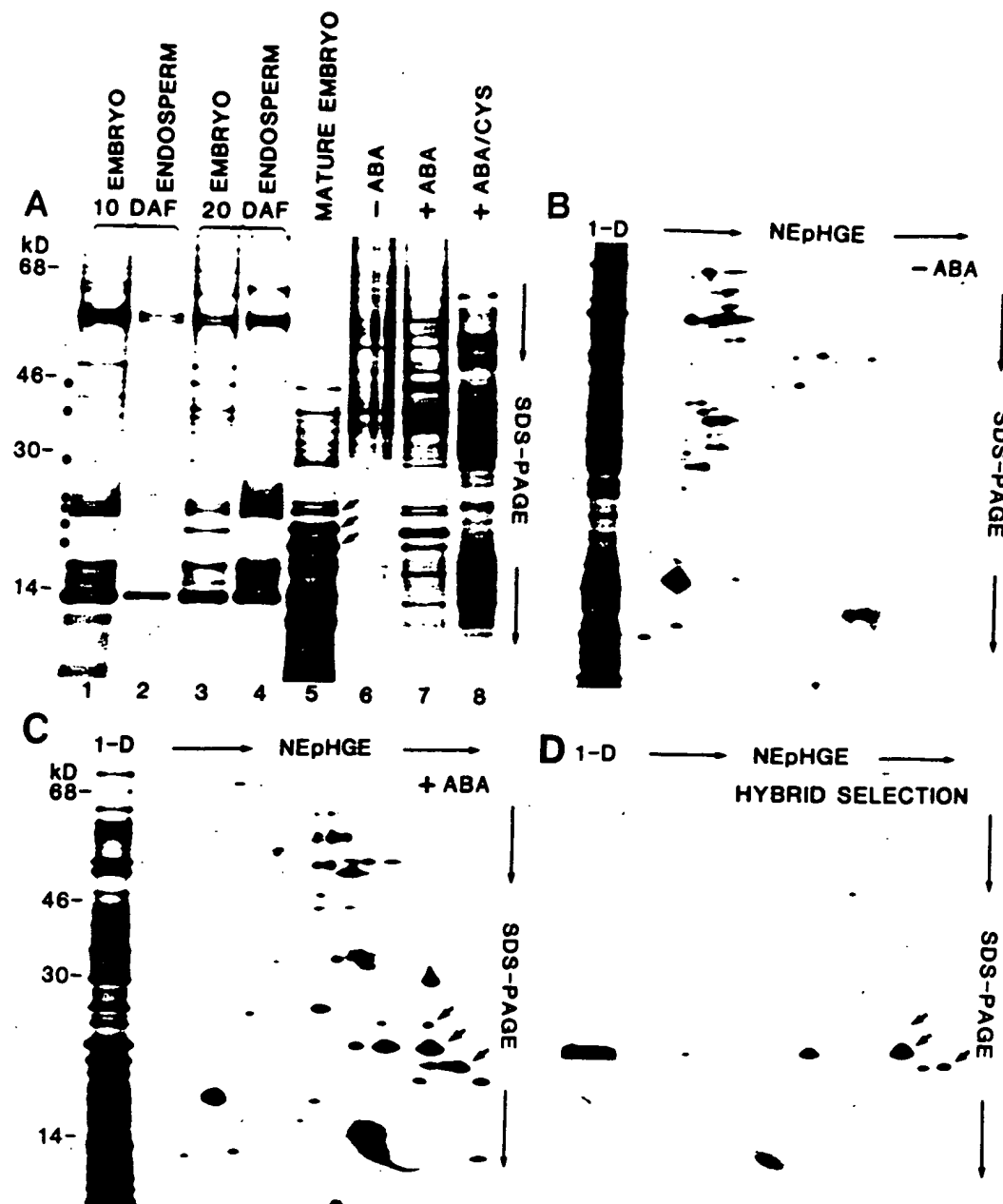


Fig. 1. Translation products of developing and germinating seed mRNAs. (A) One-dimensional SDS-PAGE of *in vitro* products of mRNAs from embryo and endosperm half-seeds harvested 10 days after flowering (DAF) (lanes 1 and 2), embryo and endosperm at 20 DAF (lanes 3 and 4), mature embryo at 30 DAF (lane 5), mature embryo germinated 3 days and then incubated 12 h without (lane 6) and with 10  $\mu$ M ABA (lanes 7 and 8). ABA-responsive products are marked on the far left with dots; 23-, 21- and 20-kD polypeptides, referred to in panels C and D, are marked with arrows (lane 5). Products in lanes 1–7 were labelled with [ $^{35}$ S]methionine, those in lane 8 with [ $^{35}$ S]cysteine. (B) Two-dimensional NEpHGE/SDS-PAGE of [ $^{35}$ S]methionine-labelled translation products of mRNAs from germinated embryo half-seeds incubated without ABA (as in A, lane 6). One-dimensional gel also shown at left. (C) NEpHGE/SDS-PAGE of translation products of mRNAs from germinated embryo half-seeds incubated with 10  $\mu$ M ABA (as in A, lane 7). (D) NEpHGE/SDS-PAGE of translation products of poly(A) RNA selected by hybridization to RAB 21 cDNA. Mol. wt markers are indicated at the left in kD.

## Results

### Isolation of an ABA-inducible cDNA encoding a 21-kD polypeptide

As a first step toward isolating genes from rice whose expression is affected by ABA, mRNA populations from developing rice seeds were analyzed by *in vitro* translation and SDS-PAGE of the protein products. This experiment identified mRNAs encoding prominent polypeptides of mol. wts 45, 39, 30, 25, 23, 21 and 20 kD which accumulate late

during rice embryogenesis (Figure 1A, lanes 1–5, see dots). These mRNAs are long-lived in mature grain harvested 30 days after flowering (DAF) (Figure 1A, lane 5) and disappear completely during normal germination (Figure 1A, lane 6). However, their accumulation can be recapitulated during germination by a 12-h incubation with ABA (Figure 1A, lane 7). These results indicate that the levels of mRNA encoding these polypeptides are modulated by ABA.

Two-dimensional separation of these translation products revealed that the 23-, 21- and 20-kD polypeptides are com-

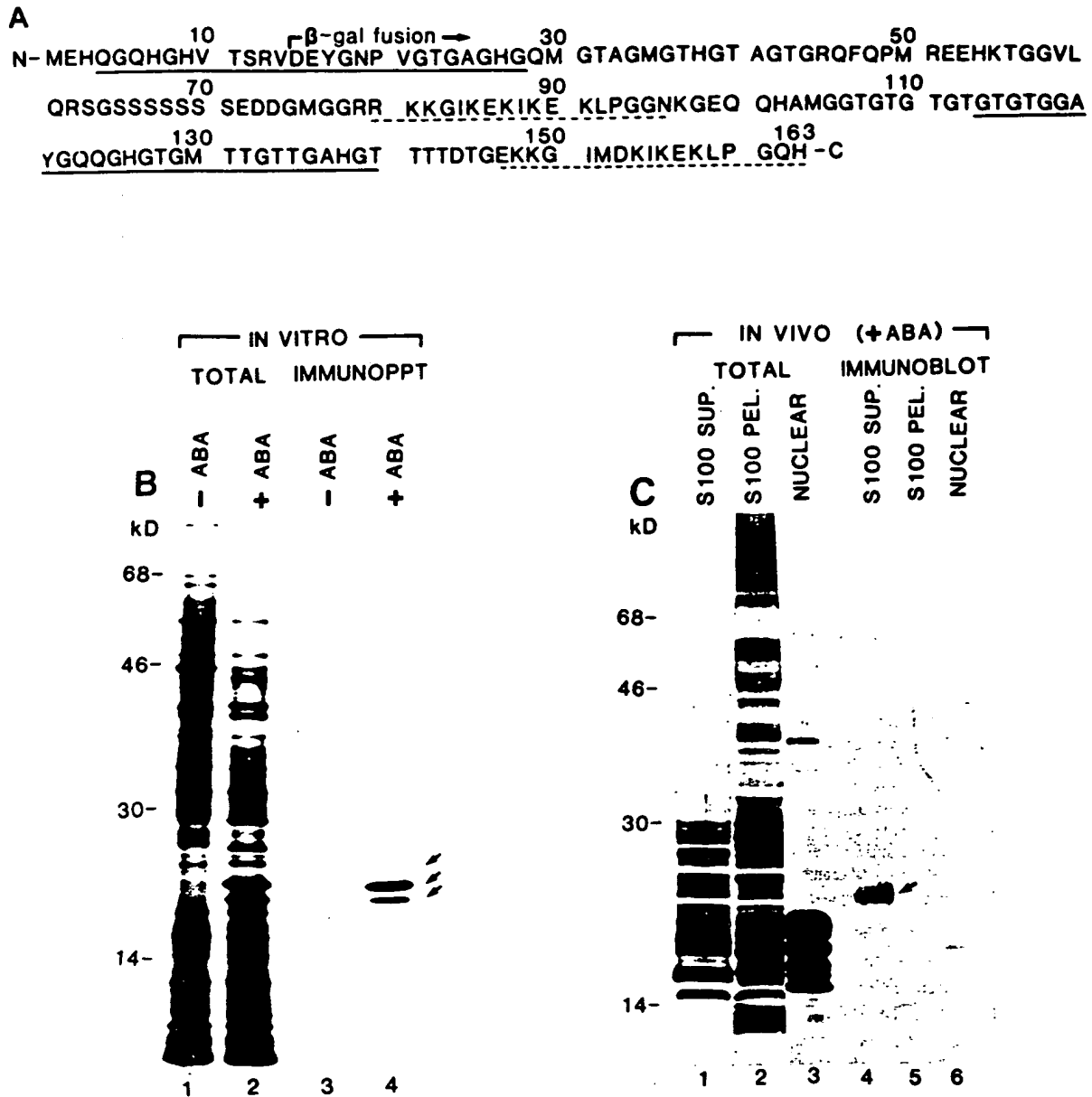


Fig. 2. Structure and expression of the RAB 21 protein. (A) Amino acid sequence of the RAB 21 protein deduced from the ORF of the cDNA (see Figure 5). The A- and B-type sequence repeats are underlined with solid or dashed lines respectively. The splice site used in constructing the  $\beta$ -galactosidase fusion proteins is marked by an arrow. (B) Immunoprecipitation of *in vitro* synthesized RAB 21 polypeptide. Total translation products of mRNAs from embryo half-seeds germinated 3 days and then incubated without (lane 1) or with 10  $\mu$ M ABA (lane 2); immunoprecipitates of these products with antibodies raised against 'in frame'  $\beta$ -galactosidase/RAB 21 fusion protein (lanes 3 and 4). (C) Immunodetection of RAB 21 protein synthesized *in vivo* in leaves sprayed with aqueous 100  $\mu$ M ABA solutions. Total proteins of cytosolic, organellar and nuclear leaf cell fractions were silver-stained (lanes 1–3), or electroblotted onto nitrocellulose and the RAB 21 protein was then detected with antibodies raised against the  $\beta$ -galactosidase/RAB 21 fusion protein (lanes 4–6). Mol. wt markers are indicated at the left in kd.

prised of multiple isoelectric forms (Figure 1C, see arrows). The even spacing of spots suggests that some of this heterogeneity may be due to serial charge differences, a common electrophoretic artefact. Alternatively, some of the different isoforms may be products of different genes (see Discussion). Differential labelling experiments suggest that the 23-, 21- and 20-kd polypeptides are rich in methionine but lacking in cysteine (Figure 1A, lanes 7 and 8).

A cDNA library was constructed using the ABA-treated seed mRNA as template. ABA-responsive cDNA clones were isolated by differential screening and subsequent Northern blot analysis. One clone, called pRAB 21, was chosen for further characterization. Figure 1D shows that pRAB

21 hybridizes to mRNA(s) that encode the prominent polypeptide(s) of apparent mol. wt 21 000. At the low hybridization stringency shown here, mRNAs encoding the 23- and 20-kd polypeptides are also hybrid-selected (the 23-kd group is very faint in this exposure). At higher stringencies of hybrid-selection, only the 21-kd polypeptides are seen (not shown). This cross-hybridization, together with several lines of evidence discussed later, indicate that the mRNAs for the 23-, 21- and 20-kd polypeptides are homologous.

**RAB 21 is a basic, glycine-rich protein that accumulates in the cytosol of ABA-treated cells**  
 The amino acid sequence of RAB 21 was determined from

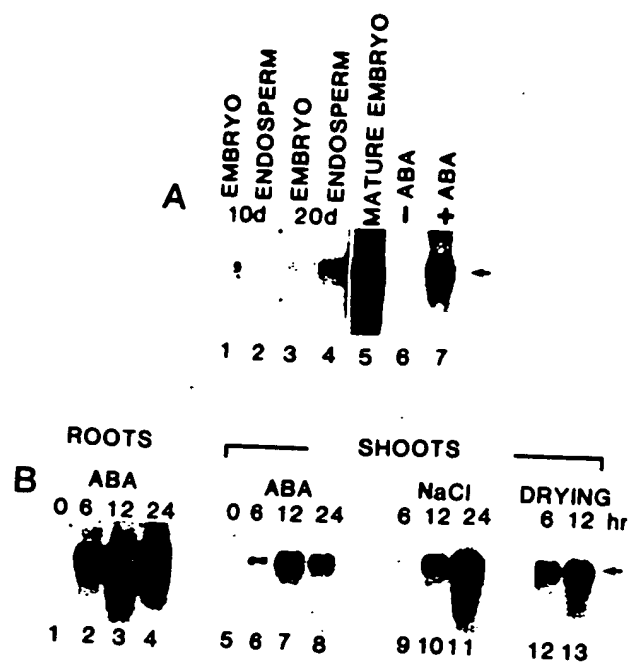


Fig. 3. (A) Steady-state RAB 21 mRNA levels in rice seeds, leaves and roots. 15  $\mu$ g of total RNA was used per lane. Northern blot analysis of RAB 21 mRNA levels in embryo and endosperm half-seeds harvested 10 DAF (lanes 1 and 2), embryo and endosperm at 20 DAF (lanes 3 and 4), mature embryo at 30 DAF (lane 5), mature embryo germinated 3 days and the incubated 12 h without (lane 6) and with 10  $\mu$ M ABA (lane 7). (B) Accumulation of RAB 21 mRNA in roots of hydroponically grown rice following addition of 10  $\mu$ M ABA to media (lanes 1-4), and in shoots of the same plants following addition of 10  $\mu$ M ABA to media (lanes 5-8), addition of NaCl (200 mM) to media (lanes 9-11), and air-drying the whole plants (lanes 12,13). Drying was accomplished by removing plants from growth media and air-drying on a laboratory bench at 27°C for the indicated times. Arrows mark the single major transcript of 850 bases.

the nucleotide sequence of the RAB 21 cDNA and genomic clones (see Figure 5). The RAB 21 open reading frame (ORF) (ATG 93-TGA 665) encodes a polypeptide of mol. wt 16 529 (Figure 2A), roughly 4500 smaller than the mol. wt predicted from the mobility of the RAB 21 protein on SDS-PAGE. However, several lines of evidence suggest that this is the correct sequence of the RAB 21 protein. First, the ORF encodes eight methionines and no cysteines, consistent with the results of *in vitro* translation experiments with labelled cysteine or methionine (Figure 1A). Second, the ORF-encoded polypeptide is basic (approximate pI = 9.4), in keeping with mobility of the RAB 21 polypeptide(s) in NEpHGE (Figure 1D). Third, the GC content of the ORF is high (70%) while that of the 5'- and 3'-untranslated regions is low (45%). This GC-codon bias has been noted in the coding regions of other seed protein genes (Rogers *et al.*, 1985). To obtain direct confirmatory evidence, antibodies were raised against portions of the RAB 21 protein. Figure 2B and C show that antibodies against a  $\beta$ -galactosidase/RAB 21 fusion protein (RAB 21 residues 15-163) specifically recognize *in vitro* and *in vivo* synthesized polypeptides which correspond to RAB 21. Furthermore, fusion proteins with the RAB 21 cDNA in all three reading frames were produced in *Escherichia coli* and analyzed on Western blots with polyclonal antibodies raised against an extract of total soluble proteins from mature rice seed. Only the 'in

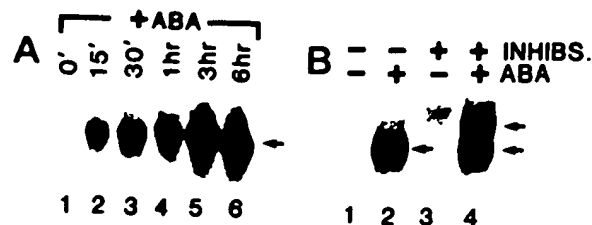


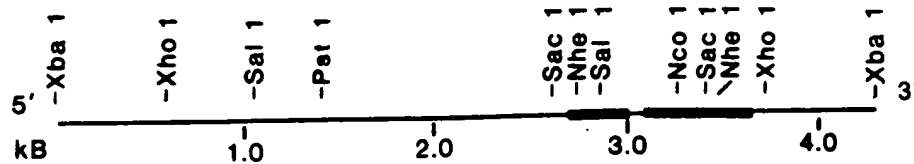
Fig. 4. Steady-state RAB 21 mRNA levels in suspension cells. (A) Time course of accumulation of RAB 21 mRNA in suspension cells incubated for 0 min, 15 min, 30 min, 1 h, 3 h and 6 h with 10  $\mu$ M ABA (lanes 1-6). (B) The effect of pre-incubation with inhibitors of protein synthesis. RAB 21 mRNA levels after 6 h incubations without (lane 1) and with 10  $\mu$ M ABA (lane 2). Lanes 3 and 4 show the same incubations following a 1.5-h pre-incubation with inhibitors of protein synthesis [cycloheximide (10  $\mu$ M), anisomycin (60  $\mu$ M) and chloramphenicol (100  $\mu$ M)]. The arrows mark the two different transcripts which accumulate. The size of the smaller transcript is 850 bases.

frame' fusion was immunoreactive (results not shown), demonstrating conclusively that the other two possible reading frames are not translated *in vivo*.

*In vitro* transcription/translation of RAB 21 cDNA constructs was used to show that the predicted stop codon (TGA, 665) terminates the RAB 21 ORF. In these experiments, pTZ19U vectors (USBC) containing either the *NheI*-*SacI* (57-670) fragment which terminates at the expected TGA, or the *NheI*-*NheI* (57-741) fragment, were used as templates to produce polypeptides *in vitro*. These two polypeptides had identical mobilities in SDS-PAGE (not shown), confirming that TGA at residue 665 is used as the stop codon.

Analysis of the RAB 21 protein sequence using Harr plots revealed that the protein contains a duplicated domain structure. Each domain contains an A and B repeat (Figure 2A). The A and B repeats are adjacent in the C-terminal domain, but are separated in the N-terminal domain by a portion of the ORF containing the gene's single intron. An algorithm which predicts the secondary structure of RAB 21 suggests that most of the polypeptide, including the A repeats, is composed of random turns due to the irregular spacing of glycine residues. The B-repeats are very basic and contain the conserved sequence KIKEKLPG which may be part of a helix bounded by turns. Further computer analysis showed weak homology between the entire protein (or the A repeats alone) to regions of several viral nuclear proteins. The highest homology occurred with the glycine-alanine copolymer domain of the major nuclear antigen encoded by the Epstein-Barr virus (Hennessy and Kieff, 1983). Not surprisingly, the very basic, lysine-rich B-repeats are weakly homologous to various DNA-binding proteins such as histones. These results suggest that the RAB 21 polypeptide may be a nuclear protein. To examine this possibility, RAB 21 antibodies were used to probe soluble, organellar, and nuclear fractions prepared from leaves of ABA-treated rice plants. These experiments show that the RAB 21 protein is found in the soluble fraction (Figure 2C). Since the protein has not been detected in crude organellar, chloroplastic, mitochondrial or nuclear extracts, we suggest that RAB 21 is a cytosolic protein. However, definitive proof of its cellular location remains to be established by immuno-electron microscopy.

**A**



**B**

CTGCAC - 150:

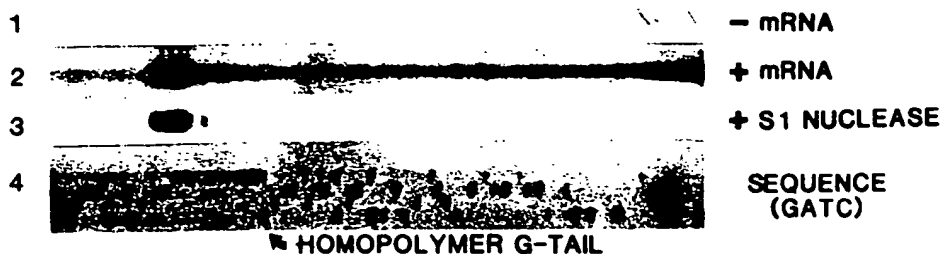
AGCAGGTGACCCCTTGTCAACACCGTCATGTACGAGGCTGCTTACCACCTGCCCTCACTGCCACCAGCGCTCTCCGCCCGGTGCAATACAAAGAACAAATC - 140:  
GAACGGTCATATAAGCTAAGACCCACTACCGATTAAACCTATCATTCCCAACAATCTAATCCACTTATTTCTCTTCCCATGATCTTATCCTCTCAITTTCTC - 130:  
CTCACTACTTTTGCATTTTGTACGAAACACAATGACACCGTCSAAGAAAGCTCGTGGAGCACCGTAGCCAGCAATCACCACAAACACAGACGGGAGGAGCTC - 120:  
CGCAGCGGCATCGGACGGCGATGACACAACCGGACGCAAGAGGAGGAGGAGCTGCCGATCATCTCGTGTGGCGGAGGAGCTACTGCCCATGC - 110:  
CAATGACAGCGGGGACGCGCAACACAAAAGGGGGAGGATGCCGGGACGACCGTAGTACCATGAACCAAGATCATGTGAACGGAGGACCGGACGAGG - 100:  
CTTGGACCTCTGCCGCCGACGTGAAGAGCGTGATCTGTAGAAAGAGATGTAGACCAGATGCCGACGCAACTTAGCCCTGCAAGTCACCCGACTGCATAT - 90:  
CGCTGCTTCCCTCGCTCATGTACACAATCAGCTTGGCTTATCTCTGCATACTTGTGCTTTGTTTCCGCTGCCCAAAATAGAACAAGACAGAGCTGGCT - 80:  
TTTGTTCGAGAGTTTACTGCTATTGTAGCGCTATTITGAATTTTGTGTACTTTATCTAATTAATCAATAAAGCTGTTTCTATTCTATTTGACTCAATC - 70:  
TTGAATCCATTGATCTCTTGGTGTGCACTCAGTATGTTGAATATTGATTCCGTTGAAACAATCTTGGTTAAGCGTTGGAACATTTTATCTGTTCCGT - 60:  
GAAACATCGTAATATTTTCTGTTGAAACAATTTTATCCGACAGCAGCGTCCAACAATTTACACCAATTTGGACGTGTGTATACATACAGCTCCCAACTC - 50:  
AAACTGACCACCGAGTTGAAGGTATACAAAGTGAACTTATTCTATTAAGAGCCGACAGATGGCGCTGCCGCTGCTGCCAACGACACCGCTTCACGC - 40:  
CCATGAGCGCTATTTATTTTTAAAAAATATTTCAACAAAAAGAGAACGGATAAAATCCATCGAAAAAATAAATTTCTACGCATCTCTCCTTATCTC - 30:  
CATCCACGGCGAGCACTCATCCAAACCGTCCATCCACGGGACAGTACACACATAGTTATCTCTCTCCCGCGATGAGTCACCAACCGCTGCTTCGA - 20:  
GAAACCGCTCCCGGACACCGTACGTGCCACCGCGCGGCTTCCGCGCTGGACACGTCCGGCTCCTCTCCCGCGCGGCTGCCGACCGCTCCACCGGCTCC - 10:  
CGCACACGTCTCGCTGTCTCCTCCACCCATCGCGCTGCCAAACGAGCTCATCTCTCGCTCTCTCCGCTTATAAATGCGGGCACCACTTCACCTGCT - 1

TGCACACCAAGCAAGAGCTAAGTGAGCTAGCCACTGATCACAAGAACCTCGATCTCTGAGAGTGTTTTTACGCTTTAGCTTAAGCAGCATCAGCA - 100  
CGACGGCGCAGCAGCGGCACGTCACCGACCGCGCTCGACGACTACCGCAACCGCTCGGCACCGCGCGCGCACCGGCAGATGGCCACCGCGCGCATCGCG - 200  
ACGCACCGCAGCGCGCGCACCGCGCGGCACTTCCAGCCGATCAGCAGGAGCACAAGACCGCGCGGCTCTGCAACCGCTCCGCGAGCTCCAGCTGAACCT - 300  
CGGACACACCTTTGACCCCAATTTCTTACCCCCCACTAAACCTTGGTACAACTTGTGAAATTTCTAATGCTCTGTGACAGTCTCAGGATGATCGA - 400  
ATCGGACGGCAGCAGCAAGACGGCATCAAGCAGAACATCAACCAAGCTTCCCGCGCGCAACAAGCGGCAGCAGCAGATGCCATGCGCGCGCACCGGCA - 500  
CGGCACCGGCACCGGCACCGGAACCGCGCGGCTACCGGCACCGCGCACCGCACCGGATGACACCGGCACCGCGCGCACCGGCACCGCACCGCAG - 600  
CAGGCACCGCGCAGCAAGACGGCATCATGCAAGATCAACCAAGCTTCCCGCGCGCAGCAGTCACTCGACACACCAACCAACCATGTGTGCTGCG - 700  
CGGCACCGCGCGCACCTCACCTTCTGAATAAAGATGAGTAGCGAGCGCAATAAAGGAAAAAATAATGTTACTGTCTGTGATGAGTGTGAATGT - 800  
GTGATGGCGTTTCCAGTACGGACCGTGTCTGCGTGTGTTGTACTGTGAAGTACGCTGTGTATGTACGTCGTATGTGTACAATTCGGTTATACTGT - 900  
TGTGAAATACGTATAAATATGTGTACATTATACCGTCTATATCTCATGTCGATATGTACACAACGTTTTCGTGATCGTTATAATGTTATTTTTTC - 1000

TTATTCGATCAATCTGGATCATAGGAGCTC

**C**

**EXTENSION PRODUCT (81b)**



**Fig. 5.** Structure of the RAB 21 cDNA and gene. (A) Restriction map of the 4.3-kb *Bam*HI genomic fragment containing the RAB 21 gene. The map was generated by multiple enzyme restriction analysis. The two *Sac*I sites are at -52, downstream of CAAT, and at +670, downstream of the TGA terminating the ORF (see Figure 2B). (B) Nucleotide sequence of the RAB 21 gene. The 2.4-kb *Pst*II-*Xho*I genomic fragment (see A) and the cDNA were sequenced as described in Materials and methods. The ORF is in bold-faced type, the single intron is in lower case. Putative transcription and the start and stop codons of the ORF are underlined. Polyadenylation sites found in the cDNAs are marked with an asterisk. Various GC-rich repeats in the proximal promoter region, including five sequences which resemble the consensus binding sites of the SP1 animal transcription factor (Briggs *et al.*, 1986) are discussed in Results. (C) Primer extension using a synthetic oligonucleotide (+84 to +108). The end-labelled oligo was used as a primer to reverse transcribe without RNA (panel 1) and with 1  $\mu$ g of poly(A)<sup>+</sup> RNA from ABA-treated half-seeds without (panel 2) and with subsequent incubation with S1 nuclease (panel 3). Panel 4 shows the 5' sequence (GATC top to bottom) of the pRAB 21 cDNA obtained with the same oligo as primer.

### *Accumulation of RAB 21 mRNA is induced in different tissues by ABA and water-stress*

Northern blot analysis was used to analyze the steady-state levels of RAB 21 mRNA in various rice tissues during different treatments. Experiments with total RNA from developing seed tissues (Figure 3A) confirmed the patterns of RAB 21 mRNA accumulation as assayed by *in vitro* translation (Figure 1A). During seed development, low levels of RAB 21 mRNA are found in both endosperm and embryo seed halves (Figure 3A, lanes 1–4). High levels of this mRNA accumulate in embryos between 20 DAF and maturity (30 DAF) and survive as long-lived mRNAs in the resting grain (Figure 3A, lane 5). Homologous signals were also seen among long-lived RNAs extracted from mature seeds of maize, barley, wheat and millet (not shown), suggesting that proteins homologous to RAB 21 are functionally conserved among the Graminae. As was shown for barley (Mundy *et al.*, 1986), mature rice endosperm does not contain detectable mRNA levels. Therefore RAB 21 mRNA was not assayed in this tissue.

The long-lived RAB 21 mRNA in rice embryos is rapidly turned over at the onset of germination (Figure 3A, lane 6). However, accumulation of this mRNA can be recapitulated by incubating the germinating embryos in 10  $\mu$ M ABA (Figure 3A, lane 7). This indicates that ABA affects the steady-state level of RAB 21 mRNA.

RAB 21 is not present at detectable levels in the roots and shoots of hydroponically-grown, 1-month-old rice plants (Figure 3B, lanes 1 and 5). However, upon addition of ABA to the solution bathing the roots, the mRNA accumulates in both organs and reaches a steady-state level after 12 h (Figure 3B, lanes 2–4, 6–8). At this time the level of RAB 21 mRNA is induced at least 20-fold over the control. Therefore the increase of the RAB 21 mRNA levels in the presence of ABA is not restricted to seed tissues.

To determine if the expression of RAB 21 could also be induced by water-stress, NaCl was added to the hydroponic solution to a final concentration of 200 mM. Figure 3B (lanes 9–11) indicates that RAB 21 mRNA levels increase in response to prolonged growth in solutions containing elevated salt concentrations. The time course of induction in response to NaCl is similar to that obtained with ABA; the maximal level of accumulation (20- to 30-fold induction) being attained after 12–24 h. The pattern of RAB 21 mRNA accumulation was unchanged when both ABA and salt were added to the hydroponic solution. RAB 21 mRNA levels are also induced by desiccation of rice plants (Figure 3B, lanes 12, 13). These results strongly suggest that the accumulation of RAB 21 mRNA is regulated by the water status of the plant.

Detectable RAB 21 mRNA accumulation in whole rice plants generally requires 3–4 h of ABA treatment. This long response time may reflect a low rate of exogenous hormone uptake by roots and translocation to the leaves. To examine this possibility, RAB 21 mRNA accumulation was measured in cultured suspension cells derived from embryogenic calli. In contrast to whole plants, the increased RAB 21 mRNA levels are easily detected in suspension culture cells after only 15 min of ABA treatment (Figure 4, lanes 1 and 2). During prolonged incubation with ABA, RAB 21 mRNA levels increase steadily to a maximum of ~20 times the control level at 3–6 h (Figure 4, lanes 3–6).

The rapidity of the ABA response in cell cultures suggests

that protein synthesis is not required for RAB 21 gene expression. To obtain direct evidence on this point, cells were treated with a combination of protein synthesis inhibitors affecting 70S and 80S ribosomes which reduced total protein synthesis by >90%. The protein synthesis inhibitors neither induce the expression of RAB 21 (Figure 4, lanes 1 and 3) nor block its induction by ABA (lanes 2 and 4). A longer transcript (1000–1200 bases) that hybridizes to the RAB 21 cDNA probe accumulates in the presence of the inhibitors. These results have also been seen in leaf tissues treated with ABA plus inhibitors (results not shown). This longer transcript has not yet been characterized but it may correspond to the primary transcript of a homologous gene which contains an intron of 300–400 bp.

### *Structure of RAB 21 gene*

The nucleotide sequence of the pRAB 21 cDNA was used to deduce the primary structure of the encoded polypeptide. The clone was also used to isolate a corresponding genomic clone, gRAB 211. Analysis of sequences upstream of the RAB 21 gene's coding region identified putative regulatory elements. The restriction map and nucleotide sequence of the cDNA and the corresponding region of the genomic clone (2.5-kb restriction fragment) are presented in Figure 5A and B. The major ORF of the transcribed region is 489 bp (93–665) encoding the 162 amino acids of RAB 21 (Figure 2A). Flanking the ORF is a 92-bp 5' leader containing three stop codons and a 261-bp 3' tail containing the putative polyadenylation sequence ATAAA 12 bp upstream of the site of poly(A) addition. The 3'-non-coding regions of six homologous cDNAs were sequenced. Five of them were identical to pRAB 21, being polyadenylated at position +926. This correlates well with the size of the RAB 21 mRNA (800 nucleotides). The sequence of the sixth clone was identical to that of the genomic DNA but contained a poly(A)<sup>+</sup> tail farther downstream at position +991. The original 824-bp cDNA was shown by primer extension to lack only 18 nucleotides from the 5'-untranslated leader (Figure 5C). The transcription start site is shown in Figure 2B (nucleotide +1).

In the genomic DNA, the ORF is interrupted by a 83-bp, AT-rich intron flanked by the consensus border sequences GT and AG. The sequence of gRAB 211 was identical to the cDNA RAB 21, indicating that this gene is transcribed *in vivo*. The proximal GC-rich promoter (–200 to –1) contains a putative TATAA box (–30) and a putative CAAT box (–62). This region contains the following four types of GC-rich repeats (see Figure 5B): (1) TGCGCCACCG at –175 and –121; (2) CGCCGCGC at –167 and –129; (3) TCCGGCTCC at –143, –108 and –37; (4) GTC-TCCCT at –93 and –85.

This region also contains five other GC-rich repeats at –200, –195, –166, –133 and –48 whose opposite strand sequence shows 80% homology to the decanucleotide (G/TG/AGGCGG/TG/AG/AC/T) binding site of the SP1 transcription factor (Briggs *et al.*, 1986).

### *Discussion*

We are interested in studying the molecular mechanism of action of plant hormones on gene expression. To begin this work we isolated a cDNA encoding a major transcript that



is inducible in rice tissues by the plant hormone ABA. This cDNA was shown by hybridization and immunoassay to encode a prominent member of a group of basic polypeptides of mol. wt 23–20 kd. We call this protein RAB 21. The different RAB polypeptides may be post-translational modifications of a single gene product or products of closely related genes. The rice genome contains at least three closely-linked (within 30 bp) genes homologous to RAB 21 (K. Yamaguchi-Shinozaki, unpublished results). These genes encode proteins of slightly different amino acid sequence that may account for the different groups of polypeptides immunoreactive to RAB 21 antibodies.

Northern blot hybridizations show that RAB 21 gene expression is not tissue specific, as shown by the accumulation of its transcript in seeds, roots, leaves and in undifferentiated suspension cells. Since there at least three other rice genes closely related to RAB 21, gene-specific probes are needed to ascertain whether these genes are differentially expressed. Physiological experiments show that RAB 21 mRNA accumulates not only in ABA-treated tissues but also in leaves, roots and suspension cells under conditions of water deficit. These results suggest that RAB 21 gene expression is dependent upon the water status of plants and that ABA may act as a signal in this response. The pathway of this response is different to that mediating the more general heat shock response (Heikkilä *et al.*, 1984), because RAB 21 mRNA is undetectable in rice tissues after heat shock for various periods of time (not shown).

Experiments with cultured cells show that accumulation of RAB 21 mRNA following ABA treatment is very rapid and that it is insensitive to inhibitors of protein synthesis. These data strongly suggest that ABA-induced gene expression does not require protein synthesis but probably involves modification of pre-existing factors, as is the case for the heat shock response (Zimarino and Wu, 1987). NaCl also induces RAB 21 mRNA rapidly in cultured cells (not shown). This response is not additive to that attributed to ABA: addition of NaCl and hormone together does not 'superinduce' mRNA accumulation at any point in the time course. These results corroborate the findings of physiological (Jones *et al.*, 1986) and genetic (Chandler *et al.*, 1988) studies which show that the response of plants to water-stress is mediated by ABA at the level of specific gene expression.

To initiate studies on the molecular mechanism of ABA action, the gene encoding RAB 21 was isolated and its nucleotide sequence determined. This is the first published genomic sequence of a strongly ABA-responsive gene. The proximal promoter region is GC-rich and contains numerous repeats detailed in Results. Another group of repeats is closely related to the GC element found in various cellular and viral genes in mammalian cells. This *cis*-acting element promotes the expression of genes by binding the *trans*-acting protein factor SP1 (Briggs *et al.*, 1986). Similar sequences have been noted in the promoter of the  $\alpha$ -subunit of  $\beta$ -conglycinin (Chen *et al.*, 1986) and in that of oat phytochrome (Hershey *et al.*, 1987), plant genes which are not known to be responsive to ABA treatment. The regulatory roles of these different GC-rich repeats remain to be established by functional assays. These experiments, now in progress in our laboratory, may elucidate the molecular mechanism by which ABA regulates gene expression and mediates the adaptation of plants to water-stress.

## Materials and methods

### Plant materials

Seeds of rice (*Oryza sativa*, var. Indica, cv. IR 36) were obtained from the International Rice Research Institute, Philippines. Plants for developmental studies were grown in soil at 27°C and a day length of 11 h. Plants flowered 10–11 weeks after planting. Seeds were then collected after 10 days (milk stage), 20 days (starchy, green pericarp) and 30 days (dry, brown pericarp, mature). Plants were also grown hydroponically in Hoagland's solution supplemented with 20 mM  $\text{NH}_4\text{NO}_3$ . Preparation and treatment of embryo-containing half-seeds at 27°C was performed as described previously for barley (Mundy *et al.*, 1986). Suspension cells derived from embryonic callus of IR 36 were grown at 25°C in Kao's medium (Kao, 1977) containing 2.5 mg/l 2,4-D and 0.2 mg/l kinetin with subculturing at 1-week intervals. Cells for mRNA isolation were subcultured 5–7 days prior to harvest.

### Protein analysis

Proteins were analyzed by NEPHGE, SDS-PAGE and Western blots according to Tingey *et al.* (1987). Isolation of mRNA, *in vitro* translation using reticulocyte lysate and immunoprecipitation with Protein A Sepharose 4B were as described previously (Mundy *et al.*, 1986). Leaf cell fractions were prepared from 7-day-old plantlets sprayed three times with 100  $\mu\text{M}$  aqueous ABA solutions during the 24 h prior to harvest. Chloroplasts and mitochondria were prepared from leaf tissue after Boutry and Chua (1985) while nuclei were prepared after Green *et al.* (1987). The National Biomedical Research Foundation Protein Sequence Databank carried in the Rockefeller University 7000/40 computer was screened for sequences related to the RAB 21 ORF with the SEARCH program of Dayhoff *et al.* (1983). Detailed comparisons of protein sequences thought to be related to the RAB 21 sequence were made with ALIGN, utilizing the mutation data matrix [250 PAMs = 6, and a gap penalty of 8 (Dayhoff *et al.*, 1983)].

### Isolation of cDNA and genomic clones

Double-stranded cDNA synthesized by the RNase H method (Gubler and Hoffman, 1983) was size-fractionated on a column of Bio-gel A50m (Bio-Rad Laboratories). Molecules of 450–4500 bp were then cloned by homopolymer GC-tailing into pEMBL 12 plasmid (Dente *et al.*, 1983). Six thousand recombinant clones, replica-transferred onto nitrocellulose filters from 96-well microtiter plates, were screened with single-stranded cDNA probes synthesized from mRNA isolated from control and ABA-treated half-seed mRNAs. Hybridizations were performed in 50% formamide, 6  $\times$  SSC, 1  $\times$  Denhardt's, 0.1% SDS, 100  $\mu\text{g}/\text{ml}$  denatured salmon sperm DNA at 42°C with denatured  $^{32}\text{P}$ -labelled DNA probes (sp. act.  $1 \times 10^8$  c.p.m./ $\mu\text{g}$  DNA, final concentration  $1 \times 10^6$  c.p.m./ml). Sixty ABA-responsive clones were then tested on Northern blots of the same RNAs. Hybridization selection after Tingey *et al.* (1987) was used to identify polypeptides encoded by specific clones. Standard protocols and conditions were used for agarose electrophoresis, hybridization in 50% formamide, DNA fragment isolation with DE-81 paper, and plasmid DNA preparation (Maniatis *et al.*, 1982).

Genomic DNA was isolated from 10-day-old etiolated leaves by CsCl centrifugation (Maniatis *et al.*, 1982). Southern hybridization of restricted DNA revealed a 4.3-kb *Xba*I fragment which hybridized strongly to the selected RAB 21 cDNA. This genomic fragment was partially purified by size fractionation in agarose gels and then cloned by insertion into *Xba*I-digested lambda ZAP (Stratagene). A total of 75 000 recombinant plaques were screened on duplicate filters and three clones containing identical 4.3-kb *Xba*I inserts were identified by hybridization to pRAB 21. Excision and recircularization of pBLUESCRIPT SK(m13) plasmid following superinfection of lambda ZAP infected cells with IR 408 helper phage (Russell *et al.*, 1986) was performed according to the manufacturer's instructions (Stratagene).

### DNA sequencing and primer extension

The pRAB 21 cDNA insert and a 2.5-kb *Pst*I–*Xho*I genomic fragment taining the RAB 21 sequences were sequenced from overlapping deletions created by *Bal*31 exonuclease (Misra, 1985). Single-stranded templates were prepared by superinfecting pEMBL containing recombinants with the IR 408 helper phage (Russell *et al.*, 1986). Sequencing reactions were performed according to Biggin *et al.* (1983) and the products separated on 6% polyacrylamide/urea gels. Inosine was used to resolve GC-compressions. More than 90% of the sequence was obtained for both strands of DNA. At least two overlapping clones were used when only one strand was sequenced.

A 24-base oligonucleotide corresponding to 5' sequences of pRAB 21 was synthesized on an Applied Biosystems model 380A DNA synthesizer.

after the manufacturer's instructions. The gel-purified oligonucleotide was used for primer extension according to Shelleness and Williams (1984).

#### Fusion protein and antibody production

The *SalI* fragment of RAB 21 was ligated into the *SalI* site of expression vectors pUR 288 (in frame) and pUR 278 and 289 (out of frame controls). Ruther and Muller-Hill, 1983). The pUR 288 fusion plasmid encodes amino acids 15–163 of the RAB 21 ORF fused to the C terminus of *E. coli*  $\beta$ -galactosidase. Fusion proteins purified from cell extracts on anti- $\beta$ -galactosidase–Sepharose columns according to the manufacturer's instructions (Promega Biotech) were used for immunization of rabbits.

#### RNA blot analysis

Total RNA was prepared by a miniprep procedure (Nagy *et al.*, 1988). RNAs were separated in formaldehyde gels, blotted onto nitrocellulose and hybridized to random-primed cDNA probes after standard protocols (Maniatis *et al.*, 1982). Replicate gels were stained with ethidium bromide to ensure that samples contained approximately equal amounts of rRNA. A commercial RNA ladder was used as size marker (BRL).

#### Acknowledgements

We thank Irene Roberson for excellent technical help and Dr Brian Keith for expert advice. J.M. is on leave from the Department of Biotechnology, Carlsberg Research Laboratory, Copenhagen, Denmark. This work was supported by a grant from the Rockefeller Foundation.

#### References

- Biggin, M.D., Gibson, T.J. and Hong, G.F. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3963–3965.
- Boutry, M. and Chua, N.-H. (1985) *EMBO J.*, **4**, 2159–2165.
- Bray, E.A. and Beachy, R.N. (1985) *Plant Physiol.*, **79**, 746–750.
- Briggs, M.R., Kadonaga, J.T., Bell, S.P. and Tjian, R. (1986) *Science*, **234**, 47–52.
- Chandler, P.M., Walker-Simmons, M., King, R.W., Crouch, M. and Close, T.J. (1988) *J. Cell. Biol., Suppl.* **12C**, Abs. #L033, 143.
- Chang, T.T., Armenta-Soto, J.L., Mao, C.X., Peiris, R. and Loresto, G.C. (1986) In *Rice Genetics*. Proceedings of the International Rice Genetics Symposium, I.R.R.I., pp. 387–398.
- Chen, Z.-L., Schuler, M.A. and Beachy, R.N. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 8560–8564.
- Dayhoff, M.O., Barker, W.C. and Hunt, L.T. (1983) *Methods Enzymol.*, **91**, 524–545.
- Dente, L., Cesarini, G. and Cortese, R. (1983) *Nucleic Acids Res.*, **11**, 1645–1655.
- Finklestein, R.R. and Crouch, M.L. (1986) *Plant Physiol.*, **81**, 907–912.
- Finklestein, R.R., Tenbarger, K.M., Shumway, J.E. and Crouch, M.L. (1985) *Plant Physiol.*, **78**, 630–636.
- Galau, G.A., Hughes, D.W. and Dure, L. (1986) *Plant Mol. Biol.*, **7**, 155–170.
- Green, P.J., Kay, S.A. and Chua, N.-H. (1987) *EMBO J.*, **6**, 2543–2549.
- Gubler, U. and Hoffman, B. (1983) *Gene*, **25**, 263–269.
- Heikila, J.J., Papp, J.E.T., Schultz, G.A. and Bewley, J.D. (1984) *Plant Physiol.*, **76**, 270–274.
- Hennessy, K. and Kieff, E. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 5665–5669.
- Henson, I.E. (1984) *Ann. Bot.*, **54**, 569–582.
- Hershey, H.P., Barker, R.F., Idler, K.B., Murray, M.G. and Quail, P.H. (1987) *Gene*, **61**, 339–348.
- Hornberg, C. and Weiler, E.W. (1984) *Nature*, **310**, 321–324.
- Jacobsen, J.V. and Beach, L.R. (1985) *Nature*, **316**, 275–277.
- Jones, H., Leigh, R.A., Tomos, A.D. and Jones, R.G.W. (1987) *Planta*, **170**, 257–262.
- Kao, K.N. (1977) *Mol. Gen. Genet.*, **150**, 225–230.
- King, R.W. (1976) *Planta*, **132**, 43–51.
- Litts, J.C., Colwell, G.W., Chakerian, R.L. and Quatrano, R.S. (1987) *Nucleic Acids Res.*, **15**, 3607–3618.
- Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Misra, T.K. (1985) *Gene*, **34**, 263–268.
- Mozer, T.J. (1980) *Cell*, **20**, 479–485.
- Mundy, J., Hejgaard, J., Hansen, A., Hallgren, L., Jorgensen, K.G. and Munck, L. (1986) *Plant Physiol.*, **81**, 630–626.
- Nagy, F., Kay, S.A. and Chua, N.-H. (1988) In *Plant Molecular Biology Manual*. Martinus Nijhoff, Dordrecht, in press.

- Raikel, N.V. and Wilkins, T.A. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 6745–6749.
- Ramagopal, S. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 94–98.
- Richardson, M., Valdez-Rodriguez, S. and Blanco-Labra, A. (1987) *Nature*, **327**, 432–434.
- Rogers, J.C., Dean, D. and Heck, G.R. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 6512–6516.
- Russel, M., Kidd, S. and Kelley, M.R. (1986) *Gene*, **45**, 333–338.
- Ruther, U. and Muller-Hill, B. (1983) *EMBO J.*, **2**, 1791–1794.
- Seshu, D.V. and Sorrells, M.E. (1986) In *Rice Genetics*. Proceedings of the International Rice Genetics Symposium, I.R.R.I., pp. 369–382.
- Shelleness, G.S. and Williams, D.L. (1984) *J. Biol. Chem.*, **259**, 9929–9935.
- Singh, N.K., Braker, C.A., Hasegawa, P.M., Handa, A.K., Buckel, S., Hermanson, M.A., Pfankuch, E., Regnier, F.E. and Bressan, R.A. (1987) *Plant Physiol.*, **85**, 529–536.
- Suzuki, Y., Kuroguchi, S., Murofushi, N., Ota, Y. and Takahashi, N. (1981) *Plant Cell Physiol.*, **22**, 1085–1093.
- Tingey, S.V., Walker, E.L. and Coruzzi, G.M. (1987) *EMBO J.*, **6**, 1–9.
- Zimmarino, V. and Wu, C. (1987) *Nature*, **327**, 727–720.

Received on March 21, 1988; revised on April 25, 1988

#### Note added in proof

These sequence data will appear in the EMBL/GenBank/DBJ Sequence Databases under the accession number Y00842.

# Subdomains of the Octopine Synthase Upstream Activating Element Direct Cell-Specific Expression in Transgenic Tobacco Plants

Halina Kononowicz,<sup>1</sup> Y. Eureka Wang, Lillian L. Habeck, and Stanton B. Gelvin<sup>2</sup>

Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907

Previous work has shown that the octopine synthase (*ocs*) gene encoded by the *Agrobacterium tumefaciens* Ti-plasmid contains an upstream activating sequence necessary for its expression in plant cells. This sequence is composed of an essential 16-bp palindrome and flanking sequences that modulate the level of expression of the *ocs* promoter in transgenic tobacco calli. In this study, we have used RNA gel blot analysis of RNA extracted from transgenic tobacco plants to show that the octopine synthase gene is not constitutively expressed in all plant tissues and organs. This tissue-specific pattern of expression is determined, to a large extent, by the 16-bp palindrome. Histochemical analysis, using an *ocs-lacZ* fusion gene, has indicated that the 16-bp palindrome directs the expression of the *ocs* promoter in specific cell types in the leaves, stems, and roots of transgenic tobacco plants. This expression is especially strong in the vascular tissue of the leaves, leaf mesophyll cells, leaf and stem guard cells, and the meristematic regions of the shoots and roots. Sequences surrounding the palindrome in the upstream activating sequence restrict the expression of the *ocs* promoter to fewer cell types, resulting in a reduced level of expression of  $\beta$ -galactosidase activity in the central vascular tissue of leaves, certain types of leaf trichomes, and the leaf primordia.

## INTRODUCTION

Many eukaryotic genes contain upstream activating sequences or enhancer elements that are essential for gene transcription. These sequences are not promoters in themselves but activate transcription from linked promoters. They are independent of their orientation relative to the promoters and can be either independent (enhancers) or dependent (upstream activating sequences) upon their position relative to these promoters. Enhancers can be tissue specific or inducible. They may also be modular, in that they may be composed of a limited number of basic sequence motifs that interact in a synergistic fashion (for a review, see Wasylyk, 1988).

Upstream activating sequences have been identified for a number of plant genes. Among the best studied of these activators is that of the cauliflower mosaic virus (CaMV) 35S promoter. The CaMV 35S activator has multiple domains that can function either independently or synergistically to activate the 35S promoter in a developmentally and tissue-specific manner (Benfey et al., 1989, 1990a, 1990b). Upstream activating sequences have also been identified among genes of the *Agrobacterium tumefaciens* Ti-plasmid encoded by T-DNA. T-DNA genes containing such elements include those

encoding enzymes involved in cytokinin biosynthesis (*tmr*, de Pater et al., 1987), agropine biosynthesis (*ags*, Bandyopadhyay et al., 1989), nopaline production (*nos*, An, 1987; Mitra and An, 1989), mannopine biosynthesis (*mas*, DiRita and Gelvin, 1987; Langridge et al., 1989; Leisner and Gelvin, 1989; Comai et al., 1990), octopine production (*ocs*, Ellis et al., 1987a, 1987b; Leisner and Gelvin, 1988, 1989; Bouchez et al., 1989; Fromm et al., 1989; Singh et al., 1989), and a gene of unknown function encoding a 780-base mRNA (Bruce et al., 1988).

Recently, we (Leisner and Gelvin, 1988, 1989) and others (Ellis et al., 1987b) have shown that the octopine synthase (*ocs*) upstream activating sequence contains a 16-bp palindrome essential for *ocs* activator function. Specific proteins can bind to this palindrome (Bouchez et al., 1989; Fromm et al., 1989; Singh et al., 1989, 1990; Tokuhisa et al., 1990) and presumably are involved in the activation of the *ocs* promoter. In transgenic tobacco calli, however, sequences surrounding this palindrome are necessary for maximal stimulation of *ocs* transcription. These additional sequences include an element with the potential to form a Z-DNA structure (Leisner and Gelvin, 1989). In this study, we have examined the influence of various DNA sequences from the *ocs* transcriptional activating element upon *ocs* promoter expression in transgenic tobacco plants. We show that the complete *ocs*

<sup>1</sup> Current address: Department of Botany and Plant Pathology, Purdue University, West Lafayette, IN 47907.

<sup>2</sup> To whom correspondence should be addressed.

activator confers tissue-specific expression upon the *ocs* promoter. Moreover, DNA sequences surrounding the 16-bp palindrome within the *ocs* activator limit the expression of an octopine synthase- $\beta$ -galactosidase fusion protein to particular cell types.

## RESULTS

### RNA Gel Blot Analysis of *ocs* RNA from Plant Tissues and Organs

In previous studies (Leisner and Gelvin, 1988, 1989), we investigated the importance of various sequences upstream of the octopine synthase gene in the activation of the *ocs* promoter. These experiments utilized calli derived from transgenic tobacco plants transformed with constructions containing the *ocs* promoter and structural gene and various fragments of the *ocs* upstream activating sequence. In tobacco calli, octopine synthase activity directed by these constructions correlated with the steady-state level of *ocs* mRNA (Leisner and Gelvin, 1988). We wished to extend these studies and investigate the tissue-specific and developmental utilization of these *ocs* activator sequences in mature tobacco plants. We, therefore, regenerated plants from these transgenic tobacco calli and assayed octopine synthase activity in various tissues and organs of these plants. Control experiments indicated, however, that certain plant tissues contained substances that inhibited the octopine synthase assay (data not shown).

Because octopine synthase activity was not a reliable indicator of *ocs* gene activity in mature transgenic tobacco plants, we performed an RNA gel blot analysis of RNA derived from these plants to examine tissue-specific patterns of *ocs* mRNA accumulation. RNA was extracted from leaves (five leaves equally distributed along the stem), stems (three segments), flowers, the root tip, and the root base. Total cellular RNA (20  $\mu$ g) was fractionated by formaldehyde-agarose gel electrophoresis, blotted onto nitrocellulose, and hybridized with a DNA fragment containing the *ocs* gene. Figure 1A shows the regions of the *ocs* transcriptional activating sequence that were present in the different plants, and Figure 1B shows the different parts of the plant that were examined for *ocs* mRNA. A total of 15 plants harboring the different constructions were analyzed to determine the amount of *ocs* RNA. Figure 1C shows RNA gel blot analyses of RNA derived from plants containing the constructions pEN1, pAlu106, and pPAL16. There was, in general, a relatively high steady-state level of *ocs* mRNA in the leaves and the root tip, a lower level of *ocs* mRNA in the stem (especially the lower stem sections) and flowers, and a very low level of *ocs* mRNA in the basal sections of the root. Plants containing the constructions pENR1, pAlTaR54, and pAlu45 showed similar patterns of *ocs* mRNA accumulation (data not shown). Hybridization of the

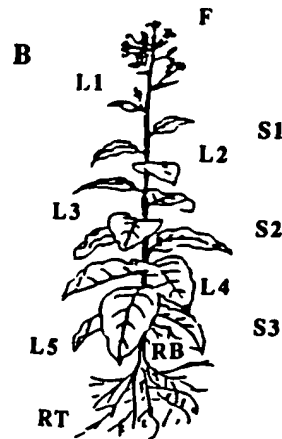
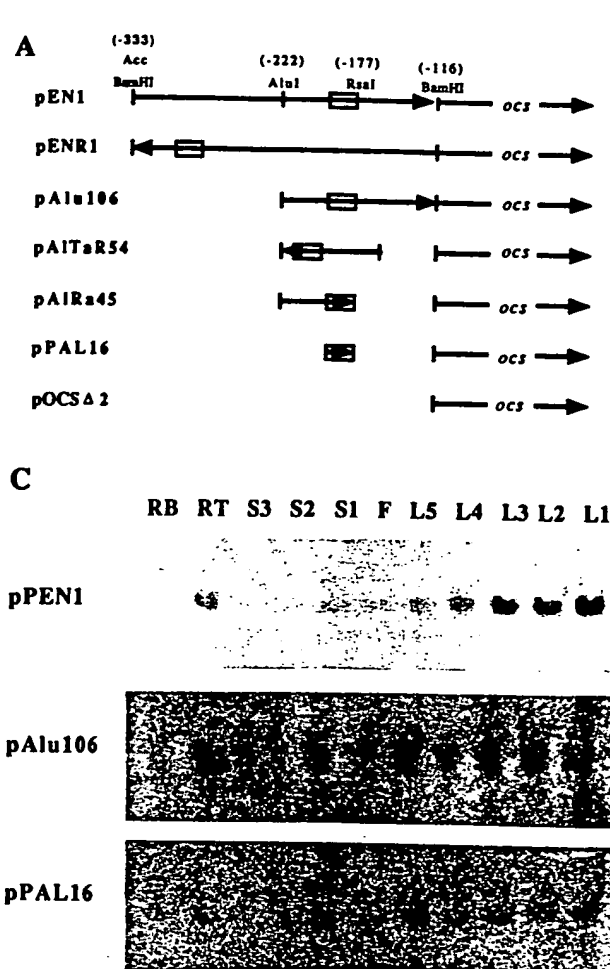
blots with an rDNA probe indicated that the lanes were equally loaded (data not shown). These data suggest that the *ocs* promoter is differentially expressed in different tissues and organs of mature tobacco plants. The data further suggest that, at the level of resolution examined in this series of experiments, different fragments of the *ocs* transcriptional activating sequence do not alter the expression pattern of *ocs* mRNA. Although the differing steady-state levels of *ocs* mRNA may also be explained by differential *ocs* mRNA stability, similar patterns of expression were observed using an *ocs::lacZ* fusion gene (see below). It is, therefore, likely that the observed steady-state levels of *ocs* mRNA resulted from tissue-specific differences in expression of the *ocs* promoter.

### Histochemical Localization of $\beta$ -Galactosidase Activity in Transgenic Plants Containing *ocs::lacZ* Fusion Constructions

To determine in which plant cell types various portions of the *ocs* transcriptional activator function, we generated transgenic tobacco plants harboring translational fusions of the *ocs* gene and the *Escherichia coli lacZ* gene. These constructions all contained the *ocs* promoter but differed in the extent of the *ocs* transcriptional activator sequences employed. The translational fusion was the same as that used by Teeri et al. (1989). The constructions used are similar to some of those shown diagrammatically in Figure 1A, except that a gene encoding an octopine synthase- $\beta$ -galactosidase fusion protein substituted for a gene encoding the complete octopine synthase protein.

The use of  $\beta$ -galactosidase activity as a histochemical marker in plant tissues depends upon the ability to distinguish between such activity encoded by the *lacZ* gene and endogenous  $\beta$ -galactosidase activity present in plant tissues. The endogenous plant activity can be eliminated by fixation of plant tissues with glutaraldehyde. We found that to abolish endogenous  $\beta$ -galactosidase activity we had to modify the protocols of Teeri et al. (1989) (see Methods). Using such modifications, we could completely eliminate the endogenous activity from plants containing the construction pOCS $\Delta$ 2 and, therefore, not expressing an *E. coli lacZ* gene (data not shown). In all further experiments, control plants containing the construction pOCS $\Delta$ 2 were examined parallel to the experimental plants. Only in those instances in which the control plants showed no endogenous activity did we examine further the experimental plants for *lacZ*-encoded  $\beta$ -galactosidase activity. For each construction, multiple tissues taken from three independent transformants were examined. There were no differences in the patterns of  $\beta$ -galactosidase expression among the individual transformants for a given construction.

Figure 2 shows sections taken through the leaves of transgenic tobacco plants harboring the constructions pEN1, pAlu106, pAlRa45, and pPAL16. The second vertical column (Figures 2B, 2E, 2H, and 2K) shows that, in all constructions,



**Figure 1.** RNA Gel Blot Analysis of *ocs* mRNA Extracted from Different Organs and Tissues of Transgenic Tobacco Plants Containing Various Portions of the *ocs* Upstream Activating Element.

(A) Schematic diagram of the portions of the *ocs* upstream activating sequence used in each construction. The arrow indicates the orientation of the upstream activating sequence fragment (left to right is the native orientation). The open box indicates the position of the 16-bp palindrome. The numbers above the map indicate the position relative to the start site of transcription (+1).

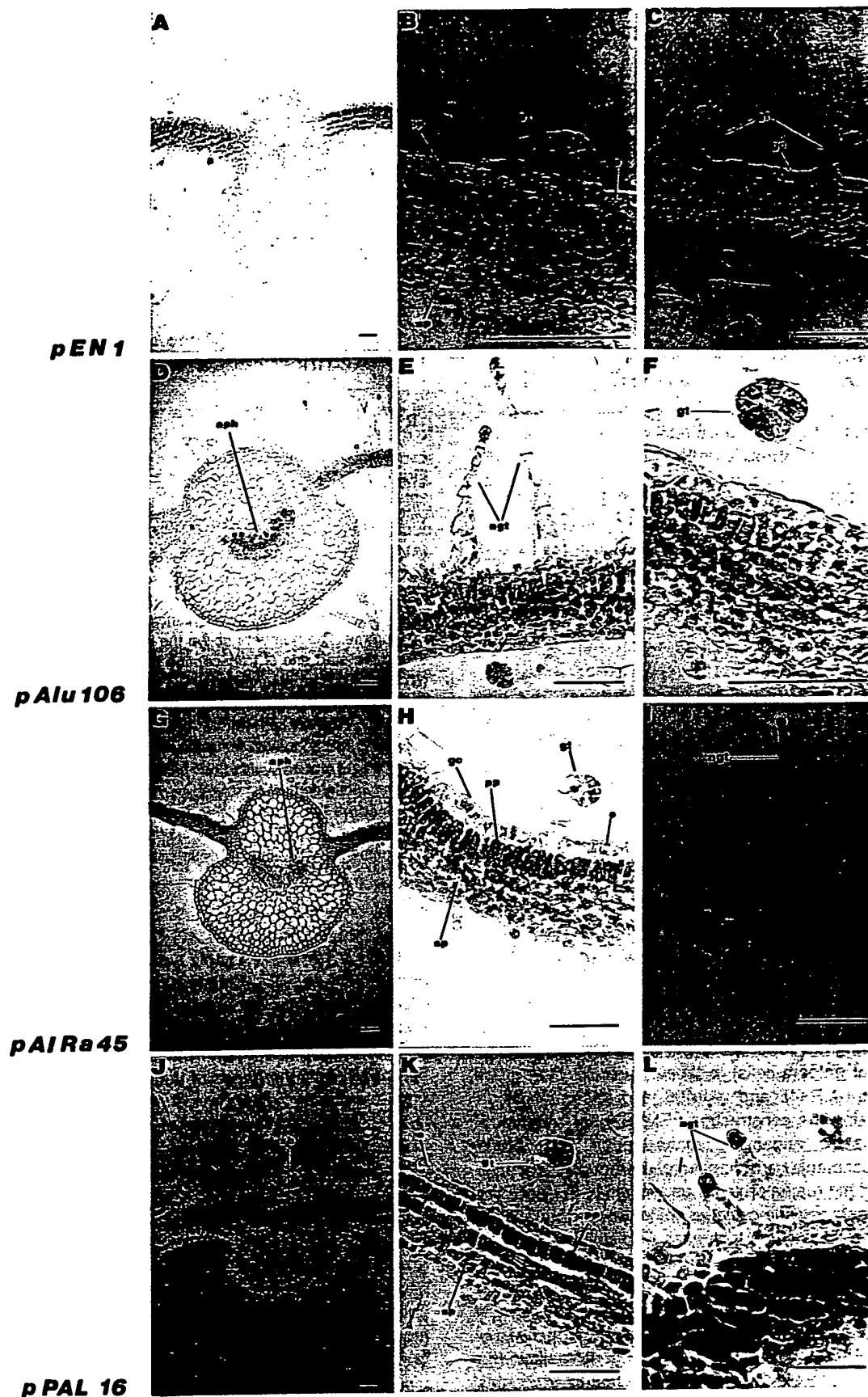
(B) Schematic representation of the tissues assayed for *ocs* mRNA. F, flower; L1 to L5, leaves from the top to the bottom of the plant, successively; RT, root tip; RB, root base; S1 to S3, stem sections from the top to the bottom of the plant, successively.

(C) RNA gel blot analysis of *ocs* mRNA from representative tobacco plants containing the constructions pEN1, pAlu106, and pPAL16. The letters above the lanes indicate the organs or tissues, as defined in (B), from which the RNA was extracted. RNA gel blot analysis was conducted using 20 µg total cellular RNA as described in Methods.

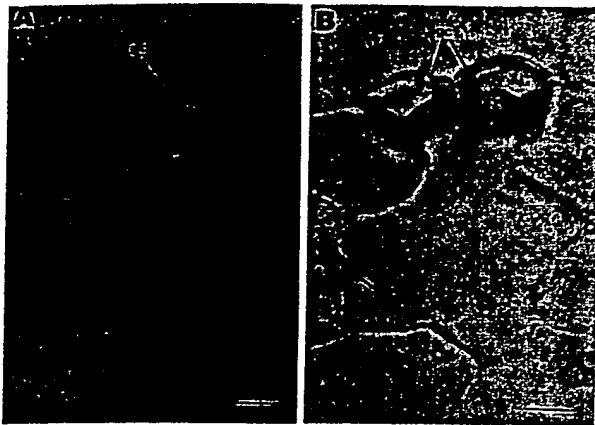
$\beta$ -galactosidase activity expressed from the *ocs::lacZ* fusion gene was present in the leaf mesophyll cells. The activity of the fusion protein in the vascular tissue was strong in those plants harboring the construction pPAL16, weak (and limited to the adaxial phloem cells) in those plants harboring the constructions pAIRa45 and pAlu106, and absent in those plants harboring the construction pEN1. Higher magnification photographs of tangential sections through the leaf blade showed that for plants harboring any of the four constructions the activity of the fusion protein was present in the palisade and spongy parenchyma cells but not in the epidermal cells or in the "stalk" cells of the trichomes. The staining of guard cells could be detected (e.g., Figures 2E, 2H, 2K, and 2L). Subtle differences in the staining patterns of the trichomes could be detected in plants containing the different *ocs* transcriptional activator constructions. Plants containing the construction pPAL16 showed intense staining in the cells in the "head" of both the glandular and nonglandular trichomes

(Figures 2K and 2L). Staining in the head of the glandular trichomes was also strong in plants containing the constructions pAlu106 and pAIRa45 but was weak in the heads of the nonglandular trichomes. In plants containing the construction pEN1, only the heads of the glandular, but not the nonglandular, trichomes showed  $\beta$ -galactosidase activity. Such lack of staining in the heads of nonglandular trichomes in pEN1 plants was highly reproducible.

Figure 3 shows the expression of the *ocs::lacZ* fusion gene in the guard cells of leaves (Figure 3A) and stems (Figure 3B). This photograph was taken of tissue from a pPAL16 plant; a similar pattern of expression of  $\beta$ -galactosidase activity in the guard cells of pEN1, pAlu106, and pAIRa45 plants was observed (data not shown). Figure 4 shows the histochemical staining of  $\beta$ -galactosidase activity in the stems of transgenic tobacco plants. The first vertical column (Figures 4A, 4E, 4I, and 4M) shows sections at the apical tip of the plant. Successive vertical columns show sections from progressively lower



**Figure 2.** Histochemical Analysis of  $\beta$ -Galactosidase Activity in the Leaves of Transgenic Tobacco Plants Harboring Various *ocs* Upstream Activating Sequence Constructions.



**Figure 3.** Histochemical Analysis of  $\beta$ -Galactosidase Activity in the Guard Cells of Transgenic Tobacco Plants Harboring the Construction pPAL16.

Tissue from transgenic plants were fixed with glutaraldehyde, stained with X-Gal, embedded, and sectioned as described in Methods. Bars = 10  $\mu$ m.

(A) Tangential section through the leaf blade of a pPAL16 plant showing guard cells. gc, guard cell.

(B) Section through the stem of a pPAL16 plant showing guard cells. gc, guard cell.

sections of the stem. Activity can be seen in the shoot apex and marginal meristem for all the constructions (Figures 4A, 4E, 4I, and 4M). Only in plants harboring the construction pPAL16, however, could activity be detected in the vascular bundle and the leaf primordium (Figure 4M). Cross-sections of stems from all four constructions showed that the *ocs::lacZ* fusion gene was active in both the inner and outer phloem cells but not in the xylem, parenchyma, or epidermal cells (Figures 4B, 4F, 4J, 4L, and 4N). In addition, the fusion gene

was active in the leaf gap areas (Figures 4C, 4D, 4G, 4K, and 4O) and in root primordia (Figures 4H and 4P) of plants containing all four constructions.

Figure 5 shows that the expression of the *ocs::lacZ* fusion gene in the roots was similar in plants harboring the different constructions. Figures 5A, 5C, 5E, and 5G show that this expression is restricted to the vascular cylinder and is absent in the epidermal and root cap cells. Cross-sections of the root (Figures 5B and 5H) indicate that  $\beta$ -galactosidase activity is detected in the central vascular cylinder but is absent in the epidermis and the cortex. Figure 5F shows that in the upper part of a root (>1 cm from the root tip) expression is limited to the vascular elements. *ocs::lacZ* gene activity in an emerging root primordium can be seen in Figure 5D.

## DISCUSSION

We have examined the organ and cellular transcriptional specificity conferred upon the octopine synthase gene promoter by various portions of the *ocs* transcriptional activating element. This element consists of a 16-bp palindromic sequence essential for *ocs* gene expression (Ellis et al., 1987a, 1987b; Leisner and Gelvin, 1988, 1989) and surrounding sequences previously shown to be important for determining the quantitative level of expression of the *ocs* gene in tobacco calli (Leisner and Gelvin, 1989). The data presented herein further demonstrate that the 16-bp palindrome is largely responsible for the tissue and cellular specificity of the *ocs* promoter. No  $\beta$ -galactosidase staining was seen in tissue sections from plants containing the construction pOCS $\Delta$ 2, indicating that the 16-bp palindrome is essential to direct *ocs* promoter activity. Sequences surrounding the palindrome, however, limit expression of the *ocs* promoter in certain cell types.

**Figure 2.** (continued).

Leaves of transgenic plants were fixed with glutaraldehyde, stained with X-Gal, embedded, and sectioned as described in Methods. Bars = 100  $\mu$ m.

(A) Cross-section through the midrib section of a leaf from a pEN1 plant.

(B) Tangential section through the leaf blade of a pEN1 plant showing a glandular trichome. e, epidermis; gt, glandular trichome; pp, palisade parenchyma; sp, spongy parenchyma.

(C) Tangential section through the leaf blade of a pEN1 plant showing both a glandular and nonglandular trichome. gc, guard cell; gt, glandular trichome; ngt, nonglandular trichome.

(D) Cross-section through the midrib section of a leaf from a pAlu106 plant. aph, adaxial phloem.

(E) Tangential section through the leaf blade of a pAlu106 plant showing a nonglandular and glandular trichome. ngt, nonglandular trichome.

(F) Tangential section through the leaf blade of a pAlu106 plant showing a glandular trichome. gt, glandular trichome.

(G) Cross-section through the midrib section of a leaf from a pAIRa45 plant. aph, adaxial phloem.

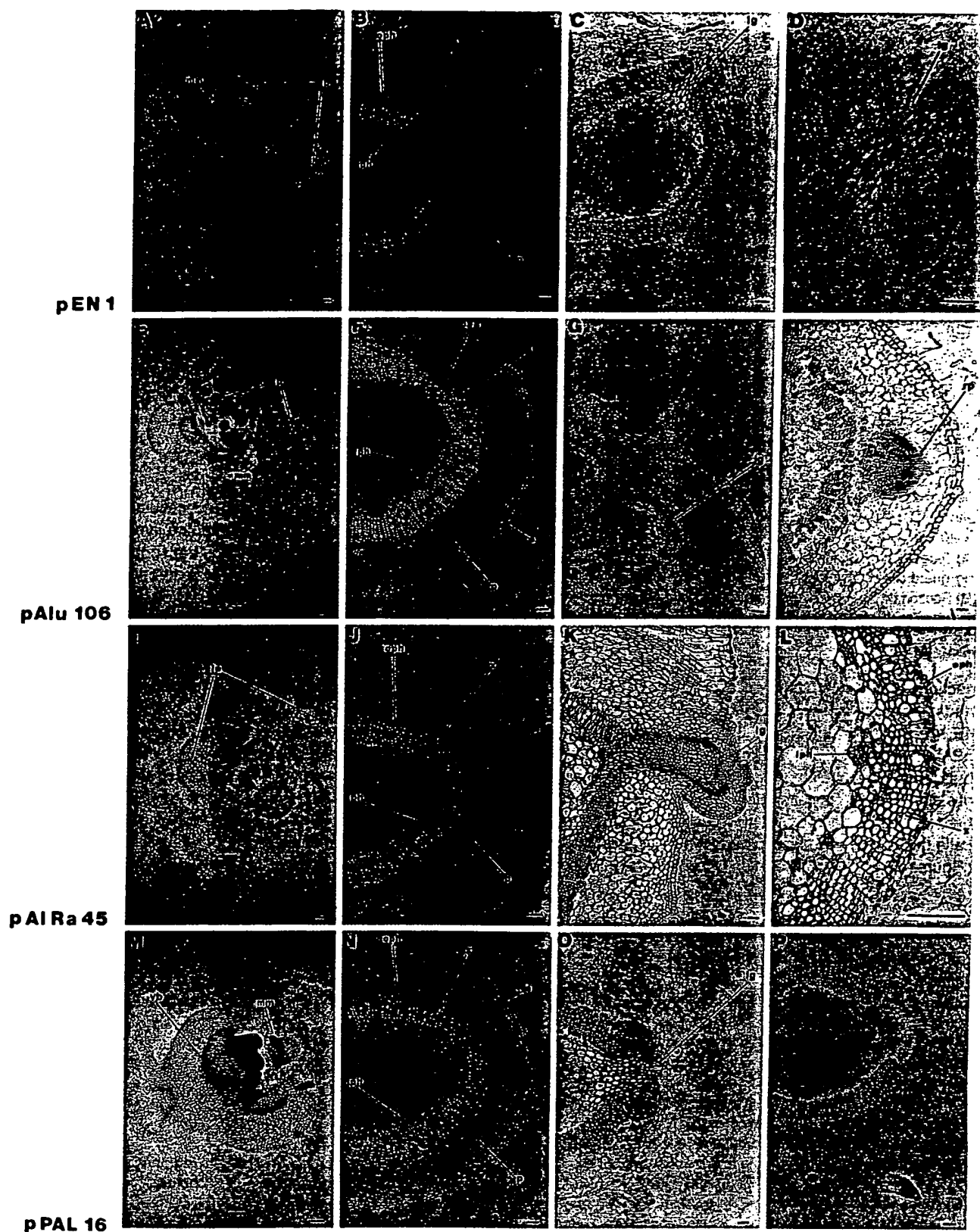
(H) Tangential section through the leaf blade of a pAIRa45 plant showing a glandular trichome. e, epidermis; gc, guard cell; gt, glandular trichome; pp, palisade parenchyma; sp, spongy parenchyma.

(I) Tangential section through the leaf blade of a pAIRa45 plant showing a nonglandular trichome. ngt, nonglandular trichome.

(J) Cross-section through the midrib section of a leaf from a pPAL16 plant. vb, vascular bundle.

(K) Tangential section through the leaf blade of a pPAL16 plant showing a glandular trichome. e, epidermis; gt, glandular trichome; pp, palisade parenchyma; sp, spongy parenchyma.

(L) Tangential section through the leaf blade of a pPAL16 plant showing nonglandular trichomes. ngt, nonglandular trichome.



**Figure 4.** Histochemical Analysis of  $\beta$ -Galactosidase Activity in the Stems of Transgenic Tobacco Plants Harboring Various *ocs* Upstream Activating Sequence Constructions.



In transgenic tobacco plants, the accumulation of *ocs* mRNA is not uniform in different tissues and organs. The expression of the *ocs* gene is greatest in the leaves and root tips, lesser in the stem and the flowers, and very low in the root base. These data suggest that the *ocs* promoter is not active to the same extent in all tobacco cells. The promoter from the nopaline synthase (*nos*) gene, as well as a number of other T-DNA-encoded genes and genes from caulimoviruses, also contain essential elements displaying a high degree of homology to the 16-bp palindrome of the *ocs* upstream activating sequence (Bouchez et al., 1989; Cooke, 1990). The *nos* promoter is not constitutive in tobacco (An et al., 1988). A gradient of activity resulted in higher expression of a *nos::CAT* fusion gene in the lower leaves and stem sections than in the upper leaves and stem sections of the plant. The expression of the *nos* promoter was also high in flower sections. Specific sequences in the *nos* promoter region were identified that were responsible for this gradient of activity (Ha and An, 1989).

Our analysis of the expression of the *ocs* gene in tobacco indicates that the *ocs* promoter has a different tissue specificity from that of the *nos* promoter. Using RNA gel blot analysis, we have shown that most of this differential tissue specificity is conferred upon the *ocs* promoter by the 16-bp palindrome portion of the *ocs* upstream activating sequence. RNA gel blot analysis was unable, however, to identify major differences in the tissue-specific pattern of *ocs* gene expression when transcription was directed by different subfragments of the *ocs* transcriptional activating element. That the tissue-specific expression of the octopine synthase gene did not appear to differ among the constructions in these experiments most likely reflects the low resolution of analysis afforded by a technique that relies upon extraction of RNA from tissues with many different cell types. DNA sequences surrounding

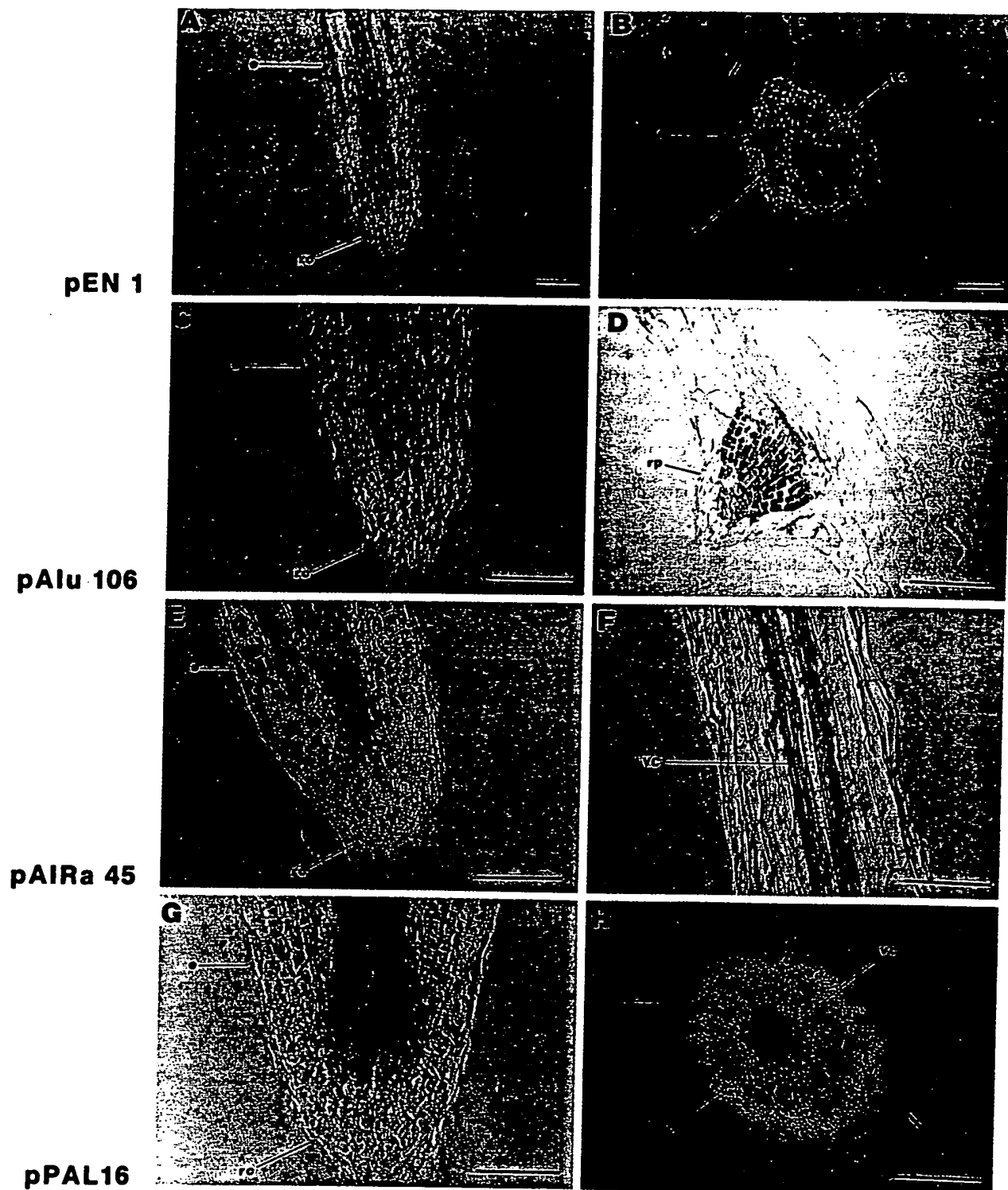
the homology found within the *ocs* 16-bp palindrome differ substantially among plant promoters, however. Because we have shown that sequences surrounding the 16-bp palindrome of the *ocs* upstream activating sequence can affect the tissue specificity of the *ocs* promoter, we speculate that the different sequences that surround the related element in other plant promoters may likewise affect the tissue specificity of these promoters.

The use of a histochemical staining technique based upon the expression of  $\beta$ -galactosidase activity directed by an *ocs::lacZ* fusion gene afforded us a much higher level of resolution of gene expression. This technique allowed us to visualize the specific cell types in which the fusion gene could be expressed. Thus, using a construction that included the entire *ocs* upstream activating sequence, we were able to detect expression of the fusion gene in the mesophyll and guard cells of the leaves, the cells at the tip of glandular trichomes, the internal and external phloem cells and guard cells of the stems, procambial and vascular cells in the root, and cells in meristematic portions of the plant. These meristematic regions included the shoot apex, leaf and root primordia, and cells in the leaf gap area of the stem. Although different constructions incorporating various subfragments of the *ocs* upstream activating sequence did not show greatly differing cellular patterns of gene expression, subtle but reproducible differences were consistently noted. These differences were most noticeable in the vascular tissues of the leaves and apical stem sections and in the leaf primordia. Expression of the *ocs::lacZ* fusion gene was undetectable in these cells in plants harboring the construction pEN1 that contains the complete *ocs* upstream activating sequence. Expression in these cells was strong in plants harboring the construction pPAL16 that contains only the 16-bp palindrome portion of the upstream activating sequence. An intermediate

Figure 4. (continued).

For each row, the left-most picture indicates a section through the apical meristem. Pictures progressively to the right in each row show sections taken from successively lower portions of the stem. Tissues were fixed with glutaraldehyde, stained with X-Gal, embedded, and sectioned as described in Methods. Bars = 100  $\mu$ m.

- (A) Section through the apical meristem region of a pEN1 plant. lp, leaf primordium; mm, marginal meristem.
- (B) Cross-section through the stem of a pEN1 plant. e, epidermis; iph, inner phloem; oph, outer phloem; p, parenchyma; x, xylem.
- (C) Cross-section through a lower portion of a stem from a pEN1 plant showing a leaf gap region. lg, leaf gap.
- (D) Higher magnification of the leaf gap region shown in (C). lg, leaf gap.
- (E) Section through the apical meristem of a pAlu106 plant. lp, leaf primordium; mm, marginal meristem.
- (F) Cross-section through the stem of a pAlu106 plant. e, epidermis; iph, inner phloem; oph, outer phloem; p, parenchyma; x, xylem.
- (G) Cross-section through a lower portion of a stem from a pAlu106 plant showing a leaf gap region. lg, leaf gap.
- (H) Cross-section through a still lower portion of a stem from a pAlu106 plant showing an emerging root. rp, root primordium.
- (I) Section through the apical meristem of a pAlRa45 plant. lp, leaf primordium.
- (J) Cross-section through the stem of a pAlRa45 plant. e, epidermis; iph, inner phloem; oph, outer phloem; p, parenchyma; x, xylem.
- (K) Cross-section through a lower portion of a stem from a pAlRa45 plant showing a leaf gap region. lg, leaf gap.
- (L) Higher magnification of a cross-section through the stem of a pAlRa45 plant. iph, inner phloem; oph, outer phloem; x, xylem.
- (M) Section through the apical meristem of a pPAL16 plant. mm, marginal meristem; vb, vascular bundle.
- (N) Cross-section through the stem of a pPAL16 plant. e, epidermis; iph, inner phloem; oph, outer phloem; p, parenchyma; x, xylem.
- (O) Cross-section through a lower portion of a stem from a pPAL16 plant showing a leaf gap region. lg, leaf gap.
- (P) Cross-section through a still lower portion of a pPAL16 stem showing an emerging root.



**Figure 5. Histochemical Analysis of  $\beta$ -Galactosidase Activity in the Roots of Transgenic Tobacco Plants Harboring Various *ocs* Upstream Activating Sequence Constructions.**

Sections were fixed with glutaraldehyde, stained with X-Gal, embedded, and sectioned as described in Methods. Bars = 100  $\mu$ m.

(A) Longitudinal section through the root tip of a pEN1 plant. e, epidermis; rc, root cap.

(B) Cross-section through the root of a pEN1 plant. c, cambium; e, epidermis; vc, vascular cylinder.

(C) Longitudinal section through the root tip of a pAlu106 plant. e, epidermis; rc, root cap.

(D) Section through the root of a pAlu106 plant showing an emerging root hair. rp, root primordium.

level of expression was seen in the vascular tissues in plants harboring the constructions pAlu106 and pAlRa45 that contain the 16-bp palindrome plus 90 or 29 bp, respectively, of DNA surrounding the palindrome. In addition,  $\beta$ -galactosidase activity was not detectable in the cells at the head of nonglandular trichomes in plants containing the construction pEN1, but expression of the *ocs::lacZ* fusion gene was readily detected in these cells from the pPAL16 plants. It thus appears that DNA sequences surrounding the 16-bp palindrome of the *ocs* upstream activating sequence limit the expression of the linked *ocs* promoter to particular cell types. Silencer sequences that restrict the expression of the *cab* gene in nonroot cells have been identified (Simpson et al., 1986). It should be emphasized, however, that the 16-bp palindrome does not render the *ocs* promoter completely constitutive; expression is still highly specific to particular cell types.

Our data appear to contradict those of Teeri et al. (1989). These authors examined the expression of an *ocs::lacZ* fusion gene in the roots of transgenic tobacco plants. Their analysis showed strong expression of  $\beta$ -galactosidase activity in the epidermal cells of these roots but no expression in the procambial or vascular cells. Our analysis indicated the opposite pattern of expression: we detected  $\beta$ -galactosidase activity only in the procambial and vascular cells but not in the epidermis or in the root cap. The *lacZ* fusion of Teeri et al. (1989) used a position within the *ocs* gene identical to ours. Although it is not possible from the article of Teeri et al. (1989) to determine which portions of the *ocs* upstream activating sequence they included in their construction, our analyses, using the entire upstream activating sequence as well as various portions of this sequence, identified a consistent pattern of expression in the roots that was not dependent upon the region of the upstream activating sequence used in the construction. It is, therefore, unlikely that the differences between our results and those of Teeri et al. (1989) can be attributed to the use of different fragments of the *ocs* upstream activating sequence. The reason for the differences in the results between the two groups cannot presently be explained.

The pattern of expression directed by the *ocs* upstream activating sequence in the roots of transgenic tobacco plants identified in this report resembles the pattern identified by Fromm et al. (1989). These authors were unable, however, to detect expression directed by this upstream activating sequence in the stem or leaves of these plants. We were easily able to detect expression in these tissues. Fromm et al. (1989)

used a construction consisting of portions of the *ocs* upstream activating sequence (including the 16-bp palindrome) fused to a truncated CaMV 35S promoter. This promoter/activator combination was affixed to a GUS gene as a transcriptional fusion. Our constructions contained various portions of the *ocs* upstream activating sequence, the *ocs* promoter, and the first third of the *ocs* structural gene. Because we used homologous portions of the *ocs* regulatory and transcribed sequences, our constructions most likely reflect the true pattern of *ocs* gene expression seen in transgenic plants. Indeed, at the level of resolution afforded by RNA gel blot analysis, our histochemical data and our analysis of the steady-state levels of RNA in the tissues of transgenic tobacco plants were consistent. It is possible that the combination of the *ocs* upstream activating sequence and the truncated CaMV promoter used by Fromm et al. (1989) resulted in a novel pattern of expression that differed from that seen using the homologous *ocs* upstream activating sequence and promoter sequences.

The 16-bp palindrome of the octopine synthase transcriptional activating sequence binds the nuclear protein OCSBF-1 (Singh et al., 1989, 1990; Tokuhisa et al., 1990). Analysis of the expression of the OCSBF-1 gene in maize indicated that the gene was most highly expressed in actively dividing and meristematic cells (Singh et al., 1990). Such an analysis has not yet been reported for tobacco. It is possible that the cell-specific expression of the *ocs* gene in tobacco reflects the prevalence of OCSBF-1 in certain cell types. The cell-specific pattern of *ocs* gene expression could then be explained by DNA sequences that regulate OCSBF-1 gene expression rather than DNA sequences that regulate *ocs* gene expression. We feel that although it is possible that this may represent a major level of *ocs* gene regulation in tobacco it cannot be the sole mechanism of *ocs* gene regulation. Sequences surrounding the 16-bp palindrome clearly modulate the expression of the *ocs* promoter both in particular cell types of tobacco plants (this study) and the quantitative level of expression of the *ocs* promoter in tobacco calli (Leisner and Gelvin, 1989). The synergistic interactions of various subdomains of the CaMV 35S promoter to yield novel cell-specific patterns of expression have been documented (Benfey et al., 1990a, 1990b). It is possible that, perhaps to a lesser extent, the *ocs* upstream activating sequence is also composed of subdomains that when deleted or combined with transcriptional activating elements from other genes can confer novel patterns of expression upon linked promoters. Experiments to test this hypothesis are currently underway in this laboratory.

Figure 5. (continued).

- (E) Longitudinal section through the root tip of a pAlRa45 plant. e, epidermis; rc, root cap.
- (F) Longitudinal section through a region of a root of a pAlRa45 plant >1 cm from the root tip. vc, vascular cylinder.
- (G) Longitudinal section through the root tip of a pPAL16 plant. e, epidermis; rc, root cap.
- (H) Cross-section through the root tip of a pPAL16 plant. c, cambium; e, epidermis; vc, vascular cylinder.

## METHODS

## Bacterial Strains and Media

*Escherichia coli* DH5 $\alpha$  was grown in LB medium (Maniatis et al., 1982) at 37°C. *Agrobacterium tumefaciens* LBA4404 and derivatives were grown in either AB minimal medium containing 0.5% sucrose or YEP medium (Lichtenstein and Draper, 1986) at 30°C. Antibiotics were used at the following concentrations ( $\mu$ g/mL): for *E. coli*—ampicillin, 100; kanamycin, 50; for *A. tumefaciens*—carbenicillin, 100; kanamycin, 100; rifampin, 10. Plasmids were mobilized from *E. coli* to *A. tumefaciens* by a triparental mating method (Ditta et al., 1980) using the mobilizing functions of pRK2013 (Figurski and Helinski, 1979).

## Plasmid Constructions

Plasmids containing *ocs::lacZ* translational fusions were derived from the previously described plasmids pEN1, pAlu106, pAIRa45, and pPAL16 (Leisner and Gelvin, 1989), which contained the octopine synthase gene and promoter and various portions of the *ocs* transcriptional activating element. pRS414 (Simons et al., 1987), containing a promoterless *lacZYA* sequence, was digested with SnaBI (which cuts in the *lacY* gene) and a HincII fragment was inserted that contains bidirectional polyadenylation signal sequences from the T-DNA 0' and 1' genes (coordinates, 21,727 to 22,440; Barker et al., 1983). The resulting plasmid (pRS414-polyA) was linearized at the unique EcoRI site preceding the *lacZ* gene. pEN1, pAlu106, pAIRa45, and pPAL16 were digested with EcoRI (which deletes the carboxy portion of the octopine synthase protein and the *ocs* polyA signal sequence) and individually ligated to pRS414-polyA. The resulting plasmids (pEN1-lac, pAlu106-lac, pAIRa45-lac, and pPAL16-lac) contain various portions of the *ocs* transcriptional activating element linked to the *ocs* promoter and an octopine synthase- $\beta$ -galactosidase translational fusion.

## Nucleic Acid Manipulations

Recombinant DNA procedures were performed according to Maniatis et al. (1982) under P1 containment conditions as specified by the guidelines of the National Institutes of Health.

RNA was extracted from plant tissue (15 plants were examined) by the following procedure: tissue (0.5–4.0 g) was frozen in liquid N<sub>2</sub> and ground to a fine powder using a mortar and pestle. One milliliter of Tris-neutralized phenol was added, and the tissue was ground further while thawing. One milliliter of RNA extraction buffer (0.5 M Tris-HCl, pH 7.0, 100 mM NaCl, 50 mM Na<sub>2</sub>-EDTA, and 1.0% SDS) was added, and the tissue was ground until it reached a homogeneous consistency. The mixture was transferred to a 15-mL polypropylene tube and centrifuged at 10,000 rpm for 15 min. The supernatant solution was removed to a fresh tube, and the pellet was reextracted with 1 mL RNA extraction buffer. The aqueous phases were combined and reextracted with 2 mL phenol. The supernatant solution was collected, and nucleic acids were precipitated by the addition of 4 mL ethanol. Following incubation at –20°C, the nucleic acids were collected by centrifugation, dried, and dissolved in 0.5 mL water; then 0.25 mL 6 M LiCl was added. The samples were incubated at 4°C overnight, and the RNA was collected by centrifugation. The RNA was dissolved in H<sub>2</sub>O and precipitated by the addition of 0.1 volume

3 M NaOAc and 2.5 volumes ethanol. The RNA was collected by centrifugation and resuspended in H<sub>2</sub>O, and the concentration was determined by UV light absorption at 260 nm.

Twenty micrograms of total cellular RNA was fractionated by formaldehyde-agarose gel electrophoresis, blotted to nitrocellulose, and hybridized to a probe containing the *ocs* gene (a 951-bp BstNI fragment; coordinates, 12,602 to 13,553; Barker et al., 1983), as previously described (Karcher et al., 1984).

Plant Transformation and Histochemical Detection of  $\beta$ -Galactosidase Activity

*Nicotiana tabacum* cv Wisconsin 38 leaf discs were infected with *A. tumefaciens* LBA4404 containing various *ocs::lacZ* fusion constructions. Kanamycin-resistant transgenic plants were regenerated according to Horsch et al. (1985). Sterile plants were maintained in Magenta cubes at 25°C and propagated vegetatively by transferring shoot cuttings. Rooted plants were assayed for  $\beta$ -galactosidase activity 4 weeks following transfer.

Histochemical staining was performed as described by Teeri et al. (1989), with the following modifications. Leaf, stem, and root tissue of transgenic and control plants were fixed with glutaraldehyde in Z buffer (pH 7.4) at room temperature. For each construction, the expression of the *ocs::lacZ* fusion gene was analyzed in a young leaf near the top of the plant and in an older leaf near the base, in the upper and lower parts of the stem, and in the roots. Leaf and stem sections were fixed in 1.0% glutaraldehyde and roots in 0.5% glutaraldehyde for 1 hr. For better penetration of the fixative, the tissue was flash frozen in liquid nitrogen and fixed again in glutaraldehyde for 1 hr. The glutaraldehyde was removed by rinsing the tissue three times in Z buffer; then the tissues were incubated in staining solution (880  $\mu$ L Z buffer, 50  $\mu$ L 100 mM K<sub>4</sub>[Fe(CN)<sub>6</sub>], 50  $\mu$ L 10 mM K<sub>4</sub>[Fe(CN)<sub>6</sub>], 8.0% X-Gal) overnight at 28°C. The material was rinsed with Z buffer twice and treated with acetomethanol (1:3) for 1 hr at room temperature to remove chlorophyll. After rinsing in 95% ethanol and absolute ethanol, the tissue was embedded in Tissue Prep 2 medium (Fisher; melting point 56.5°C) and cut into 20- $\mu$ m sections.

Control plant tissue (a kanamycin-resistant transgenic tobacco plant containing the construction pOCS $\Delta$ 2) was always prepared parallel to the experimental samples. Experimental samples were analyzed only if the control plant showed no endogenous  $\beta$ -galactosidase activity.

## ACKNOWLEDGMENTS

The authors thank Dr. Mary-Alice Webb for useful discussions and Drs. John Hamer and Susan Karcher for critical reading of this manuscript. This work was supported by a grant from the Midwest Plant Biotechnology Consortium.

Received August 19, 1991; accepted November 7, 1991.

## REFERENCES

- An, G. (1987). A potential Z-DNA-forming sequence is an essential upstream element of a plant promoter. *BioEssays* 7, 211–214.

- An, G., Costa, M.A., Mitra, A., Ha, S.-B., and Marton, L. (1988). Organ-specific and developmental regulation of the nopaline synthase promoter in transgenic tobacco plants. *Plant Physiol.* 88, 547-552.
- Bandyopadhyay, R.S., Bruce, W.B., and Gurley, W.B. (1989). Regulatory elements within the agropine synthase promoter of T-DNA. *J. Biol. Chem.* 264, 19399-19406.
- Barker, R.F., Idler, K.B., Thompson, D.V., and Kemp, J.D. (1983). Nucleotide sequence of the T-DNA region from the *Agrobacterium tumefaciens* octopine Ti-plasmid pTi15955. *Plant Mol. Biol.* 2, 335-350.
- Benfey, P.N., Ren, L., and Chua, N.-H. (1989). The CaMV 35S enhancer contains at least two domains which can confer different developmental and tissue-specific expression patterns. *EMBO J.* 8, 2195-2202.
- Benfey, P.N., Ren, L., and Chua, N.-H. (1990a). Tissue-specific expression from CaMV 35S enhancer subdomains in early stages of plant development. *EMBO J.* 9, 1677-1684.
- Benfey, P.N., Ren, L., and Chua, N.-H. (1990b). Combinatorial and synergistic properties of CaMV 35S enhancer subdomains. *EMBO J.* 9, 1685-1696.
- Bouchez, D., Tokuhisa, J.G., Llewellyn, D.J., Dennis, E.S., and Ellis, J.G. (1989). The ocs-element is a component of the promoters of several T-DNA and plant viral genes. *EMBO J.* 8, 4197-4204.
- Bruce, W.B., Bandyopadhyay, R., and Gurley, W.B. (1988). An enhancer-like element present in the promoter of a T-DNA gene from the Ti plasmid of *Agrobacterium tumefaciens*. *Proc. Natl. Acad. Sci. USA* 85, 4310-4314.
- Comal, L., Moran, P., and Maslyar, D. (1990). Novel and useful properties of a chimeric plant promoter combining CaMV 35S and MAS elements. *Plant Mol. Biol.* 15, 373-381.
- Cooke, R. (1990). The figwort mosaic virus gene VI promoter region contains a sequence highly homologous to the octopine synthase (ocs) enhancer element. *Plant Mol. Biol.* 15, 181-182.
- de Pater, B.S., Klinkhamer, M.P., Ames, P.A., de Kam, R.J., Memelink, J., Hoge, J.H., and Schilperoort, R.A. (1987). Plant expression signals of the *Agrobacterium* T-cyt gene. *Nucl. Acids Res.* 15, 8267-8281.
- DIRita, V.J., and Gelvin, S.B. (1987). Deletion analysis of the mannopine synthase gene promoter in sunflower crown gall tumors and *Agrobacterium tumefaciens*. *Mol. Gen. Genet.* 207, 233-241.
- Ditta, G., Stanfield, S., Corbin, D., and Helinski, D.R. (1980). Broad host range DNA cloning system for Gram-negative bacteria: Construction of a gene bank of *Rhizobium meliloti*. *Proc. Natl. Acad. Sci. USA* 77, 7347-7351.
- Ellis, J.G., Llewellyn, D.J., Dennis, E.S., and Peacock, W.J. (1987a). Maize *Adh-1* promoter sequences control anaerobic regulation: Addition of upstream promoter elements from constitutive genes is necessary for expression in tobacco. *EMBO J.* 6, 11-16.
- Ellis, J.G., Llewellyn, D.J., Walker, J.C., Dennis, E.S., and Peacock, W.J. (1987b). The ocs element: A 16 base pair palindrome essential for activity of the octopine synthase enhancer. *EMBO J.* 6, 3203-3208.
- Figurski, D.H., and Helinski, D.R. (1979). Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in *trans*. *Proc. Natl. Acad. Sci. USA* 76, 1648-1652.
- Fromm, H., Katagiri, F., and Chua, N.-H. (1989). An octopine synthase enhancer element directs tissue-specific expression and binds ASF-1, a factor from tobacco nuclear extracts. *Plant Cell* 1, 977-984.
- Ha, S.-B., and An, G. (1989). Cis-acting regulatory elements controlling temporal and organ-specific activity of nopaline synthase promoter. *Nucl. Acids Res.* 17, 215-223.
- Horsch, R.B., Fry, J.E., Hoffmann, N.L., Eichholtz, D., Rogers, S.G., and Fraley, R.T. (1985). A simple and general method for transferring genes into plants. *Science* 227, 1229-1231.
- Karcher, S.J., DIRita, V.J., and Gelvin, S.B. (1984). Transcript analysis of T<sub>0</sub>DNA in octopine-type crown gall tumors. *Mol. Gen. Genet.* 194, 159-165.
- Langridge, W.H.R., Fitzgerald, K.J., Koncz, C., Schell, J., and Szalay, A.A. (1989). Dual promoter of *Agrobacterium tumefaciens* mannopine synthase gene is regulated by plant growth hormones. *Proc. Natl. Acad. Sci. USA* 86, 3219-3223.
- Leisner, S.M., and Gelvin, S.B. (1988). Structure of the octopine synthase upstream activator sequence. *Proc. Natl. Acad. Sci. USA* 85, 2553-2557.
- Leisner, S.M., and Gelvin, S.B. (1989). Multiple domains exist within the upstream activator sequence of the octopine synthase gene. *Plant Cell* 1, 925-936.
- Lichtenstein, C., and Draper, J. (1986). Genetic engineering of plants. In *DNA Cloning, A Practical Approach*, Vol. 2, D.M. Glover, ed (Washington, DC: IRL Press), pp. 67-119.
- Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982). *Molecular Cloning: A Laboratory Manual*. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory).
- Mitra, A., and An, G. (1989). Three distinct regulatory elements comprise the upstream promoter region of the nopaline synthase gene. *Mol. Gen. Genet.* 215, 294-299.
- Simons, R.W., Houman, F., and Kleckner, N. (1987). Improved single and multicopy *lac*-based cloning vectors for protein and operon fusions. *Gene* 53, 85-96.
- Simpson, J., Schell, J., Van Montagu, M., and Herrera-Estrella, L. (1986). Light-inducible and tissue-specific pea *lhcp* gene expression involves an upstream element combining enhancer- and silencer-like properties. *Nature* 323, 551-554.
- Singh, K., Dennis, E.S., Ellis, J.G., Llewellyn, D.J., Tokuhisa, J.G., Wahleithner, J.A., and Peacock, W.J. (1990). OCSBF-1, a maize ocs enhancer binding factor: Isolation and expression during development. *Plant Cell* 2, 891-903.
- Singh, K., Tokuhisa, J.G., Dennis, E.S., and Peacock, W.J. (1989). Saturation mutagenesis of the octopine synthase enhancer: Correlation of mutant phenotypes with binding of a nuclear protein factor. *Proc. Natl. Acad. Sci. USA* 86, 3733-3737.
- Teeri, T.H., Lehtvasalho, H., Franck, M., Uotila, J., Helno, P., Palva, E.T., Van Montagu, M., and Herrera-Estrella, L. (1989). Gene fusions to *lacZ* reveal new expression patterns of chimeric genes in transgenic plants. *EMBO J.* 8, 343-350.
- Tokuhisa, J.G., Singh, K., Dennis, E.S., and Peacock, W.J. (1990). A DNA-binding protein factor recognizes two binding domains within the octopine synthase enhancer element. *Plant Cell* 2, 215-224.
- Waslyk, B. (1988). Transcription elements and factors of RNA polymerase B promoters of higher eukaryotes. *CRC Rev.* 23, 77-120.

nesquehonite may form directly from the resultant solutions. Therefore, the observed mineralogy and morphology of the nesquehonite and current knowledge of nesquehonite stability are both consistent with direct formation of nesquehonite under Antarctic conditions. The formation of additional nesquehonite during curation of the meteorite can be understood as continued outward migration and evaporation of the saline solutions formed in Antarctica.

Antarctic meteorites have been on the earth for on the order of  $10^4$  to  $10^6$  years and have apparently been transported great distances by glaciers (25, 26). Whether chemical weathering of meteorites occurred during their encasement in ice or only after they were exposed to the atmosphere on stranding surfaces has been uncertain (5). Results for nesquehonite from LEW 85320 suggest that, at least for salt formation, weathering may be sufficiently rapid that most observable effects can develop in tens of years rather than over thousands of years. Rapid formation of nesquehonite suggests, but does not prove, that most weathering phenomena occur after exhumation of the meteorites from deep-glacial ice. In the simplest interpretation, ages of magnesium carbonate weathering products on Antarctic meteorites would indicate the times elapsed since surface exposure of the meteorites. The amount of rust on a surface alone cannot be used as a reliable indicator of the state of preservation of meteorites. The correlation between degrees of weathering and terrestrial-residence ages might evade detection if reliance is placed solely on the rust index.

#### REFERENCES AND NOTES

1. H. Yabuki, A. Okada, M. Shima, *Sci. Pap. Inst. Phys. Chem. Res. Jpn.* 70, 22 (1976).
2. U. B. Marvin, *Antarct. J. U.S.* 15, 54 (1980).
3. J. L. Gooding, *Proc. Lunar Planet. Sci. Conf.* 12, 1105 (1981); *Geochim. Cosmochim. Acta* 50, 2215 (1986).
4. M. A. Velbel, *Meteoritics*, in press.
5. J. L. Gooding, in *International Workshop on Antarctic Meteorites*, Lunar and Planetary Institute Technical Report, 86-01, J. O. Annexstad, L. Schultz, H. Wanke, Eds. (Lunar and Planetary Institute, Houston, 1986), p. 48.
6. K. Nishizumi, *ibid.*, p. 71.
7. L. Schultz, *ibid.*, p. 80.
8. J. L. Gooding, S. A. Wentworth, M. E. Zolensky, *Geochim. Cosmochim. Acta* 52, 909 (1988).
9. U. B. Marvin and K. Morylewski, *Lunar Planet. Sci. XI*, 669 (abstr.) (1980).
10. J. L. Gooding, *Antarct. Meteor. Newsl.* 9, 22 (1986); P. Englert, personal communication.
11. D. W. Ming and W. T. Franklin, *Soil Sci. Soc. Am. J.* 49, 1303 (1985).
12. P. J. Slora, A. J. T. Jull, L. J. Toolin, *Radiocarbon* 29, 303 (1987).
13. A. J. T. Jull, D. J. Donahue, T. W. Linick, L. J. Toolin, *ibid.* 28, 191 (1986).
14. T. W. Linick, A. J. T. Jull, L. J. Toolin, D. J. Donahue, *ibid.*, p. 522.
15. T. W. Linick, thesis, University of California, San Diego (1976).
16. I. Levin et al., *Radiocarbon* 27, 1 (1985).

17. A. J. T. Jull and D. J. Donahue, *Geochim. Cosmochim. Acta* 52, 1309 (1988).
18. T. W. Linick, in preparation.
19. M. Freyre, *Cosmochim. Acta* 52, 1309 (1988).
20. W. G. McEwen, *Earth Planet. Sci. Lett.* 100, 1 (1982).
21. G. Faure, *Geochim. Cosmochim. Acta* 52, 1309 (1988).
22. S. Epstein, *Geochim. Cosmochim. Acta* 52, 1309 (1988).
23. M. M. Gr. Pillinger, *Lunar Planet. Sci. XIX*, 409 (abstr.) (1988).
24. L. Schultz, *Meteoritics* 21, 505 (1986).
25. W. A. Cassidy, E. Olsen, K. Yanai, *Science* 198, 727 (1977).

26. E. L. Fireman and T. Norris, *Proc. Lunar Planet. Sci. Conf.* 12, 1019 (1981).
27. Joint committee on powder diffraction standards, *Diffraction File* (Information Data, Swarthmore, PA).

I from the meteorite work, assistance of L. J. Toolin, D. Ming, and S. Wom. This work was supported by NSF grant EAR85-12761 to P. E. Damon and D. J. Donahue.

16 June 1988; accepted 2 September 1988

CHRISTOPHER E. FLICK

## Herbicide Resistance in Transgenic Plants Expressing a Bacterial Detoxification Gene

DAVID M. STALKER,\* KEVIN E. MCBRIDE, LORRAINE D. MALY

The herbicide bromoxynil (3,5-dibromo-4-hydroxybenzonitrile) is a photosynthetic (photosystem II) inhibitor in plants. A gene, *bxn*, encoding a specific nitrilase that converts bromoxynil to its primary metabolite 3,5-dibromo-4-hydroxybenzoic acid, was cloned from the natural soil bacterium *Klebsiella ozaenae*. For expression in plants, the *bxn* gene was placed under control of a light-regulated tissue-specific promoter, the ribulose biphosphate carboxylase small subunit. Transfer of this chimeric gene and expression of a bromoxynil-specific nitrilase in leaves of transgenic tobacco plants conferred resistance to high levels of a commercial formulation of bromoxynil. The results presented indicate a successful approach to obtain herbicide resistance by introducing a novel catabolic detoxification gene in plants.

ADVANCES IN GENETIC ENGINEERING of plants have allowed the development and transfer of agronomically important traits such as viral resistance, insect resistance, and herbicide resistance. These traits allow transfer of single dominant genes that exhibit a rapidly discernible phenotype. Herbicide resistance can be achieved by at least three different mechanisms: overproduction of a herbicide-sensitive biochemical target; structural alteration of a biochemical target, resulting in reduced herbicide affinity; or detoxification-degradation of the herbicide before it reaches the biochemical target inside the plant cell. Resistance obtained by the first two mechanisms has been developed for the herbicides glyphosate (1; 2), atrazine (3), the sulfonylureas (4), and phosphinothricin (5). A report describing the transfer to plants of a *Streptomyces* gene encoding a phosphinothricin acetyltransferase (6) resulted in phosphinothricin-resistant plants, establishing detoxification by conjugation as a viable strategy. Two advantages of a detoxification-degradation mechanism, as opposed to altering a biochemical target, are that specialized compartmentation of the detoxifying activity is not required and that greater herbicide resistance can be achieved with lower levels of detoxifying enzyme. Disad-

vantages include the potential toxicity of one or more metabolites and the possibility that detoxifying activities might react with endogenous plant compounds to impair plant function.

We have been interested in the nitrilase-containing broadleaf herbicide bromoxynil (3,5-dibromo-4-hydroxybenzonitrile), a potent photosystem II (PSII) inhibitor. Although the actual chloroplast-localized biochemical target is not well defined, there is evidence that bromoxynil acts by binding a component of the quinone-binding protein complex of PSII, inhibiting electron transfer (7, 8). It has further been suggested that a low-affinity binding site within this complex exists in the 32-kD polypeptide (9, 10). Bromoxynil has a very short half-life in the environment, as microbial populations and tolerant plant species can convert the cyano moiety of bromoxynil to the corresponding amide and acid derivatives (11-13). A natural soil isolate, *Klebsiella ozaenae*, has been identified that transforms bromoxynil to 3,5-dibromo-4-hydroxybenzoic acid, releasing ammonia (14). This reaction (Fig. 1) is carried out via a bromoxynil-specific nitrilase. The nitrilase gene (*bxn*), which is plas-

Calgene, 1920 Fifth Street, Davis, CA 95616.

\*To whom correspondence should be addressed.





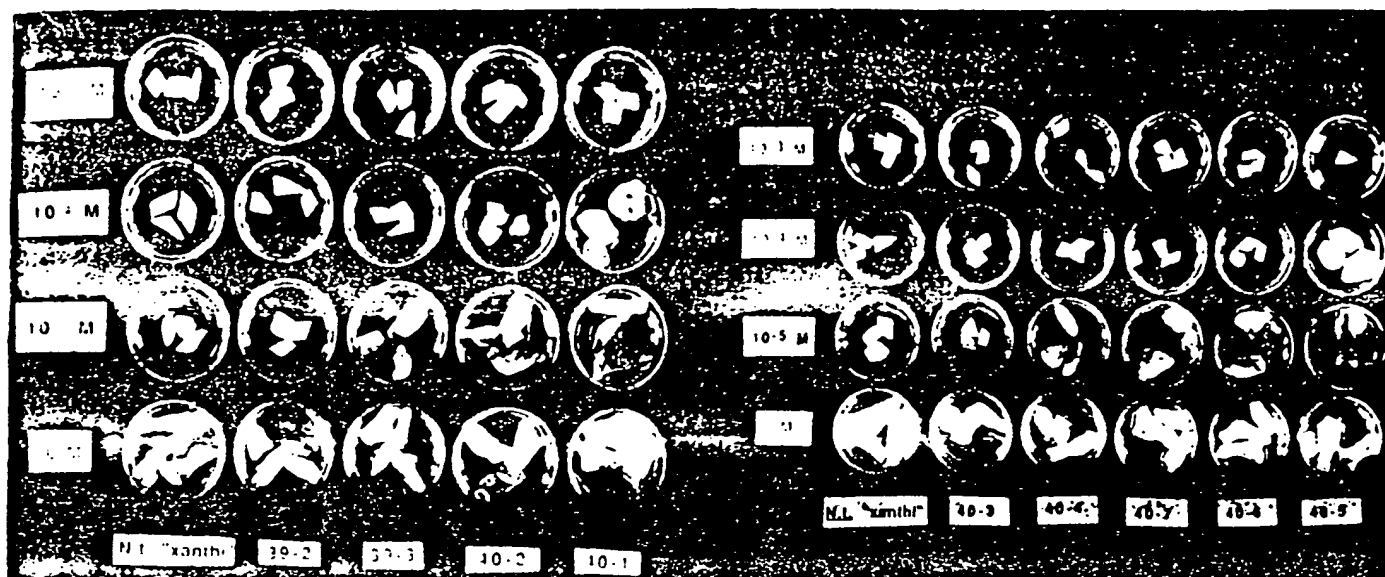


Fig. 2. Leaf sections of transgenic tobacco plants. Leaf cuttings were obtained from selected transformant plants, surface-sterilized, placed in the concentrations of bromoxynil indicated, and maintained photoautotrophically.

Chlorosis was monitored. Photographs were taken 11 days after initiation of the experiment. N.L. 'Xanthi' refers to control leaf sections.

genetic analysis by spraying with a commercial formulation of bromoxynil (Buctril, Rhône-Poulenc) at 0.5 lb (0.23 kg) per acre (normally a high field rate). Another set of plants was subjected to Buctril spraying at 0.5 lb per acre increments to determine maximum level of tolerance before chlorotic lesions appeared (Table 1). The bromoxynil resistance trait segregated as a single genetic locus in progeny of plants 39-3, 40-6, and 40-7, and as two independent loci in progeny of 40-1 in both open pollination and backcross experiments. The maximum Buctril resistance levels exhibited by the progeny of these four plants also correlated with the protein level observed in leaf tissue obtained from the respective primary  $T_1$  transformant. Progeny from 39-3, which contained the lowest level of nitrilase, exhibited symptoms at lower rates (1.5 to 2.0 lb per acre), whereas progeny of 40-1, which contained high nitrilase levels, exhibited no symptoms even at 4.0 lb per acre. This result suggests that enzyme levels determined in the primary transformants are maintained in their progeny and that the greater the level of nitrilase in the leaves of transgenic plants, the higher the level of bromoxynil resistance. Selected plants from the progeny  $T_2$  of plant 40-1 that were sprayed with bromoxynil at increasing rates are shown in Fig. 4. These transgenic plants expressing nitrilase grow, flower, and set seed normally, even when sprayed with concentrations of bromoxynil eightfold higher than the highest field rate normally used. The enzyme appears to be localized in the plant cell cytosol, as chloroplasts prepared from leaf tissue contain no detectable nitrilase polypeptide nor can the bromoxynil-specific ni-

Fig. 3. (A) Plant DNA was prepared as described (28) from 1-g leaf samples and digested to completion with Hind III, and 5 to 7  $\mu$ g of total plant DNA were electrophoresed in a 1% agarose gel. The DNA was transferred to nitrocellulose, and the blot was hybridized with a 1.2-kb Bam HI-Eco RI *bxn* gene fragment labeled with  $^{32}$ P (29). N.L. refers to DNA isolated from nontransformed tobacco tissue. Control contains the isolated 4.5-kb Hind III fragment (two copies per genome equivalent) added to DNA isolated from nontransformed plant. (B) Immunological detection of the enzyme from leaf tissue of selected transformants. Blotting procedure is described in (30-33). Control indicates nontransformed leaf tissue. The second lane contains 200 ng of purified nitrilase (16) spiked into a nontransformed leaf sample and immunoprecipitated.

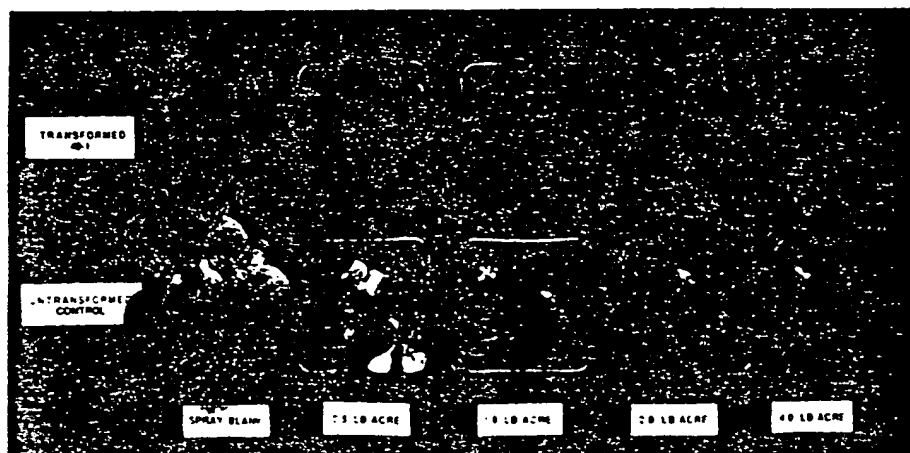
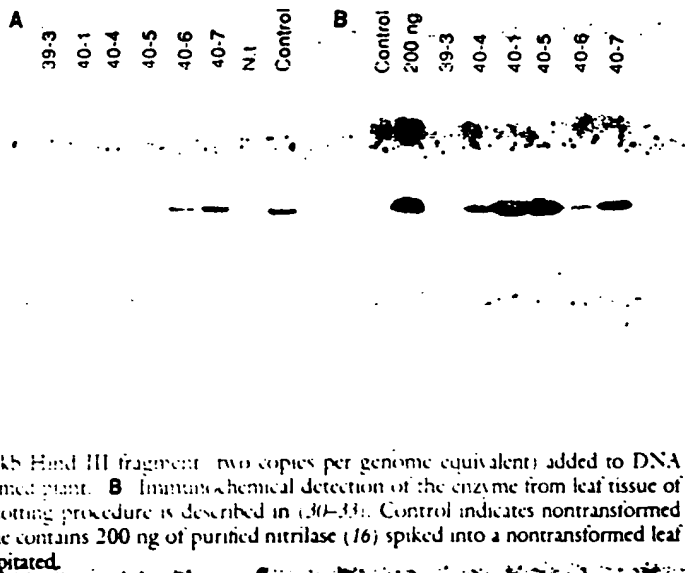


Fig. 4. Phenotype of transformed plants.  $T_2$  seed from primary transformant 40-1 was grown in 10-cm pots (three plants per pot) to the three to four leaf stage and sprayed with a commercial formulation of bromoxynil (Buctril) at the amounts indicated. Spray blank contained the formulation without the active ingredient. The photograph was taken 6 days after spraying.



release be incorporated into chloroplasts *in vitro*.<sup>24</sup>

Expression of the *8-oxo-2-oxo-bromoxynil*-specific nitrilase in photosynthetic tissue of transgenic plants confers high-level resistance to bromoxynil. The herbicide is converted to a nontoxic metabolite, 3,5-dibromo-4-hydroxybenzoic acid, by the enzyme. A low level of nitrilase expression (approximately 0.0007% total leaf protein) is required to obtain high commercial resistance levels; the higher the expression level of nitrilase, the greater the degree of resistance that can be achieved (although this is not a direct linear relation). Expression of the enzyme in plants can be achieved under the control of a tissue-specific light-regulated promoter to attain whole plant resistance. Protection of the herbicide-sensitive photosystem II target (in a phase I event) can be obtained by expressing the bromoxynil-specific nitrilase in a separate cellular compartment (cytosol). The trait is dominant and heritable, and the enzyme levels obtained in the primary transformants relate to bromoxynil resistance levels observed in the succeeding generation. These observations suggest that in the case of the bromoxynil-specific nitrilase and the phosphinothricin acetyltransferase (6), detoxification-degradation is superior to altering a biochemical target to obtain herbicide resistance in plants by gene transfer technology. Complex pathways and mutant proteins and genes do not have to be engineered nor are specialized targeting functions for specific cellular compartments needed. The sole requirement consists of a single-step conversion of the herbicide to a nontoxic metabolite.

Although extensive analysis of residues in these transgenic plants has not been completed, it will be interesting to determine the fate of the metabolite (3,5-dibromo-4-hydroxybenzoic acid) produced by the action of nitrilase. If the bromine atoms of 3,5-dibromo-4-hydroxybenzoic acid are hydrolyzed by an endogenous plant dehalogenase, then the product of this reaction would be *p*-hydroxybenzoic acid, a normal constituent of plant metabolism. In this case, no metabolites of bromoxynil would be detected. Alternatively, 3,5-dibromo-4-hydroxybenzoic acid could be concentrated in the vacuole or transported out of the cell or both may occur. The bromoxynil-specific nitrilase is active *in vitro* only as a dimer (16), suggesting that this polypeptide must undergo formation in the plant cell into an active complex. Supporting this phenomenon is the appearance of a faint undenatured band at approximately 72-kD on immunoblots (25). An analogous result was observed on simultaneous expression of the bacterial *luxA* and *luxB* genes in transgenic plants

for detection of inactive luciferase. These results indicate that bacterial proteins can form active enzymes in a diverse plant cell environment.

The gene encoding the bromoxynil-specific nitrilase can be used as a selective marker for rapid screening at the whole plant level. Leaf sections can be excised from the plant and screened photoautotrophically, or individual leaves can be treated with commercial bromoxynil-formulations without harm to the remainder of the plant. At high bromoxynil concentrations, symptoms can be scored within 24 to 36 hours after application. Finally, expression of the *K. ozonae* bromoxynil-specific nitrilase in transgenic plants to obtain a bromoxynil-resistant phenotype is another example that agronomically important germplasm can be derived from a microbial source and that novel and diverse biochemical pathways can be introduced into plants without deleterious effects.

## REFERENCES AND NOTES

1. L. Comar *et al.*, *Nature* 317, 741 (1985).
2. D. M. Shah *et al.*, *Science* 233, 478 (1986).
3. A. Y. Cheung, L. Bogorad, M. Van Montagu, J. Schell, *Proc. Natl. Acad. Sci. U.S.A.* 85, 391 (1988).
4. R. S. Chalk and T. B. Ray, *Nature* 223, 1148 (1984).
5. G. Donn, E. Fischer, J. A. Smith, H. M. Goodman, *J. Mol. Appl. Genet.* 2, 621 (1984).
6. M. DeBlock *et al.*, *EMBO J.* 6, 2513 (1987). The commercial formulation of phosphinothricin used was BASTA (Hoechst).
7. J. S. Van Rensburg *et al.*, *Plant J.* 54, 515 (1982).
8. G. F. Sanders and K. E. Pallett, *Proc. Biochem. Soc.* 26, 116 (1986).
9. W. F. J. Vermaas, K. E. Stenback, C. J. Arntzen, *Anal. Biochem.* 139, 231, 226 (1984).
10. Z. S. Zilberstein, L. G. Paley, *Plant Growth Res.* 3, 347 (1982).
11. R. F. Collins *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 70, 181 (1973).
12. A. E. Smith, *Proc. Natl. Acad. Sci. U.S.A.* 71, 276 (1974).
13. J. L. Buckland *et al.*, *Protein Sci.* 4, 149 (1973).
14. K. E. McBride, J. W. Kenny, D. M. Stalker, *Appl. Environ. Microbiol.* 52, 325 (1986).
15. D. M. Stalker and K. E. McBride, *J. Bacteriol.* 169, 955 (1987).
16. D. M. Stalker, L. D. Mahj, K. E. McBride, *J. Biol. Chem.* 265, 6310 (1988).
17. J. K. O'Neal, A. R. Pokalsky, K. L. Kiehne, C. K. Shewmaker, *Nucleic Acids Res.* 15, 8661 (1987).
18. E. M. Tobin, *Proc. Natl. Acad. Sci. U.S.A.* 75, 4749 (1978).
19. S. L. Berry-Lowe, T. D. McKnight, D. M. Shah, R. B. Mcgaher, *J. Mol. Appl. Genet.* 1, 483 (1982).
20. G. Ooms *et al.*, *Plant J.* 15 (1982).
21. Construction of a chimeric small subunit-barnase gene for expression in plants. The nucleotide sequence of a 1212 bp Pst I-Hinc II DNA segment encoding the bromoxynil-specific nitrilase has been described (16). This sequence contains 65 bp of 5' untranslated nucleotides. To facilitate removal of a portion of these excess nucleotides, plasmid pBrx9 (15) was digested with Pst I and treated with nuclease Bal 31 and Bam HI linkers added to the resulting ends. Bam HI-Hinc II fragments containing a functional *bxn* gene were cloned into the Bam HI-Sma I sites of pCGN365 (a pUC18 derivative containing a chloramphenicol resistance marker). A plasmid, pBrx25, was generated that contained only 11 bp of 5' untranslated bacterial sequence. A Bam HI-Sac I DNA fragment containing the *bxn* gene was excised from pBrx25 and cloned to the Bam HI-Sac I sites of pCGN1510 to create plasmid
22. S. J. Dellaporta, J. Wood, J. R. Hicks, *Plant J. Biol. Rep.* 1, 19 (1983).
23. T. Maniatis, E. F. Fritsch, J. Sambrook, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982.
24. For each sample, 1 g fresh weight of leaf material was ground to a powder in liquid nitrogen. Polyvinylpyrrolidone (0.5 g) was added, then 2 ml of an extraction buffer containing 0.1M potassium phosphate, pH 6.8, 0.15M NaCl, 10 mM EDTA, bovine serum albumin (25 mg/ml), 0.3% Tween 20, 0.05% NP-40, 10 mM dithiothreitol, 10 mM thioacetamide, 1 mM phenylmethylsulfonyl fluoride, and 10 mM  $\beta$ -mercaptoethanol. Samples were centrifuged at 15,000g for 15 min. 15  $\mu$ l of nitrilase antiserum obtained from rabbits and 125  $\mu$ l of 10% (w/v) suspension of cleaned Sephadex *se-4b* (Calbiochem) were added to each supernatant. After 1 hour incubation at room temperature, samples were centrifuged and the pellet was washed twice with 3 ml of a solution containing 50 mM tris, pH 7.5, 1 mM EDTA, 0.5% NaCl, and 0.05% NP-40. The pellets were resuspended in 30  $\mu$ l of a solution containing 125 mM of tris, pH 6.8, 4% SDS, 20% glycerol, 10%  $\beta$ -mercaptoethanol, and 0.05% bromophenol blue. The samples were then heated for 5 min at 95°C, spun in the microfuge for 5 min, and the supernatants electrophoresed in 10% acrylamide gels (22). The resolved polypeptides were transferred to nitrocellulose filters as described (32). Filters were incubated in Blotto (33) overnight at 4°C or for 1 hour at room temperature and then incubated for 30 min at room temperature in Blotto containing a 1:200 dilution of antiserum to nitrilase. Filters were washed for 10 min in 20 mM tris, pH 7.5, and 150 mM NaCl and then for 20 min in the same buffer containing 0.05% Tween 20, and for another 10 min in buffer without Tween 20. Then, 50 mM Blotto containing 25  $\mu$ Ci of  $^{125}$ I-labeled protein A was added to the filters and incubated at room temperature with agitation for 2 hours. The filters were washed 2 hours in 50 mM tris, pH 7.5, 1M NaCl, and 0.5% Sarkosyl, and then 1 to 2 hours in 50 mM tris, pH 7.5, 5 mM EDTA, 150 mM NaCl, 0.5% Triton X-100, and 0.1% SDS. After rinsing and drying, filters were exposed to Kodak AR X-ray film.
25. U. K. Laemmli, *Nature* 227, 680 (1970).
26. W. N. Burnette, *Anal. Biochem.* 112, 195 (1981).
27. D. A. Johnson, I. Gutsch, J. Sportsman, J. Elder, *Cann. J. Bot.* 71, 38 (1983).
28. The authors thank Dr. R. Boone-Poulenc Agre for

# Molecular Cloning of a Cysteine Proteinase Inhibitor of Rice (Oryzacystatin)

Claim 63  
Oryzacystatin

HOMOLOGY WITH ANIMAL CYSTATINS AND TRANSIENT EXPRESSION IN THE RIPENING PROCESS OF RICE SEEDS\*

(Received for publication, June 2, 1987)

Keiko Abe, Yasufumi Emori†, Hiroto Kondo, Koichi Suzuki†, and Soichi Arai

From the Department of Agricultural Chemistry, The University of Tokyo, Yayoi-cho and †the Department of Molecular Biology, The Tokyo Metropolitan Institute of Medical Science, Honkomagome, Bunkyo-ku, Tokyo 113, Japan

A cDNA clone for a cysteine proteinase inhibitor of rice (oryzacystatin) was isolated from a  $\lambda$ gt10 cDNA library of rice immature seeds by screening with synthesized oligonucleotide probes based on partial amino acid sequences of oryzacystatin. A nearly full-length cDNA clone was obtained which encoded 102-amino acid residues. The amino acid sequence of oryzacystatin deduced from the cDNA sequence was significantly homologous to those of mammalian cystatins, especially family 2 cystatins. Oryzacystatin contained the sequence Gln-Val-Val-Ala-Gly conserved among most members of the cystatin superfamily. The gene for oryzacystatin was transcribed into a single mRNA species of about 700 nucleotides. The content of mRNA reached its highest level 2 weeks after flowering and then gradually decreased to undetectable levels at 10 weeks. This feature of transient expression is coordinate with that of glutelin (a major storage protein), although the expression of oryzacystatin precedes that of glutelin by about 1 week.

Proteinaceous cysteine proteinase inhibitors of animal origin have been extensively investigated to clarify their chemical structures and physiological functions. These inhibitors are considered to have evolved from a cognate ancestral gene and form the cystatin superfamily which is further classified into three different families on the basis of their molecular structures (1). Family 1 cystatins lack disulfide bonds; human cystatins A (2) and B (3) and rat cystatins  $\alpha$  (4) and  $\beta$  (5) are typical examples. Family 2 cystatins contain two disulfide bonds as exemplified by human cystatin C (6), human cystatin S (7), bovine colostrum cystatin (8), and chicken cystatin (9). Cystatins of both families are also characterized by their molecular weights that range from 10,000 to 20,000. Family 3 cystatins comprise kininogen segments (10). The structures of various cystatins of these three families have been elucidated by amino acid sequencing or by cDNA cloning. All the elucidated cystatins conserve the sequence Gln-Val-Val-Ala-Gly or its homologues (11) which might be involved in the inhibition of cysteine proteinases (EC 3.4.22).

However, no detailed information was available about plant cystatins. We purified a proteinaceous cysteine proteinase

inhibitor from rice seeds (12) and characterized it as a cystatin (13). In the present paper we name this protein oryzacystatin.

It would be interesting to compare the amino acid sequence of oryzacystatin to those of animal cystatins in order to find out if any homology exists. The comparison may also provide information useful to clarify the structure-function relationships of oryzacystatin.

For this purpose we cloned and analyzed cDNA encoding oryzacystatin. We then analyzed the change in the amounts of corresponding mRNA during the ripening of rice seeds. The change in the amount of the mRNA encoding rice glutelin was also investigated in order to understand the physiological function of this inhibitor.

## EXPERIMENTAL PROCEDURES<sup>1</sup>

**Materials**—*Escherichia coli* DNA ligase and amino myeloblastosis virus reverse transcriptase were obtained from P-L Biochemicals and Bio-Rad, respectively. DNA polymerase I from *E. coli*, its Klenow fragment, terminal deoxynucleotidyltransferase, bacterial alkaline phosphatase, and T4 polynucleotide kinase were purchased from Takara Shuzo Co. The restriction enzymes used were products of Takara Shuzo Co., Toyobo Co., and P-L Biochemicals. The oligo(dT)-cellulose was a product of P-L Biochemicals. A nick translation kit and [ $\gamma$ -<sup>32</sup>P]ATP (3000 Ci/mmol) were obtained from Amersham Corp. [ $\alpha$ -<sup>32</sup>P]dCTP (3000 Ci/mmol) was obtained from Du Pont-New England Nuclear.

**Rice Seeds**—The rice cultivar Nihonbare (*Oryza sativa* L. japonica), a middle ripening variety, was used. The plant was grown in the Experimental Farm at the University of Tokyo, and the seeds were harvested at 1-week intervals after flowering. The seeds were aseptically removed from the pods, frozen in liquid nitrogen, and stored at -80 °C until used.

**Determination of Partial Amino Acid Sequence and Designing of Oligonucleotide Probes**—Oryzacystatin was purified from matured rice seeds as described (12). The purified oryzacystatin was digested with metalloendopeptidase (EC 3.4.24) (*Grifola frondosa*, Seikagaku Kogyo Co.) (14) and the digests were separated on a Vydac ODS column (for details see Miniprint). Two sets of oligodeoxynucleotide mixed probes (A and B) were used. Probe A was based on the partial amino acid sequence of peptide fraction 6 and probe B on that of peptide fragment 1 (Table 1S). They were synthesized with a DNA synthesizer (Applied Biosystem Inc., model 380B) and labeled at their 5' ends with [ $\gamma$ -<sup>32</sup>P]ATP and T4 polynucleotide kinase in a buffer containing 25 mM Tris-HCl (pH 9.0), 5 mM MgCl<sub>2</sub>, 5 mM dithiothreitol, and 25  $\mu$ g/ml bovine serum albumin. Probes A and B are a 16 mix of 17-mers and

\* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) J03469.

<sup>1</sup> Portions of this paper (including part of "Experimental Procedures," part of "Results," Figs. 1S-4S, and Table 1S) are presented in miniprint at the end of this paper. Miniprint is easily read with the aid of a standard magnifying glass. Full size photocopies are available from the Journal of Biological Chemistry, 9650 Rockville Pike, Bethesda, MD 20814. Request Document No. 87M-1837, cite the authors, and include a check or money order for \$2.80 per set of photocopies. Full size photocopies are also included in the microfilm edition of the Journal that is available from Waverly Press.

a 1024 mix of 17-mers, respectively, their sequences being shown in Fig. 2S.

**Construction of a cDNA Library**—Total RNA was extracted by the phenol-SDS<sup>2</sup> method (15) from rice seeds harvested at 2 weeks after flowering. Poly(A<sup>+</sup>) RNA was purified by oligo(dT)-cellulose column chromatography (16). After the synthesis of single-stranded cDNA by reverse transcriptase with total poly(A<sup>+</sup>) RNA as a template, double-stranded cDNA was synthesized essentially according to the method of Gubler and Hoffman (17) using *E. coli* polymerase I. Flush ends of the cDNA were generated with T4 DNA polymerase, and the double-stranded cDNA was treated with *Eco*RI methylase. Phosphorylated *Eco*RI linker (12-mer) was ligated, digested with *Eco*RI, and inserted into the phage vector  $\lambda$ gt10. The DNA was then packaged into bacteriophage particles using the packaging extract GP10-P (Stratagene) and grown on *E. coli* strain C600Hfl. The yield was  $5 \times 10^5$  plaques/ $\mu$ g of mRNA.

**Isolation of cDNA Clones for Oryzacystatin**—Recombinant plaques were transferred onto nylon filters (Hybond N, Amersham Corp.). The filters were fixed with ultraviolet light and prehybridized for 20 h in 50 mM Tris-HCl (pH 8.0) containing 1 M NaCl, 10 mM EDTA, 0.2% Ficoll 400, 0.1% SDS, 200  $\mu$ g/ml heat-denatured *E. coli* DNA and 200  $\mu$ g/ml salmon sperm DNA at 49 °C for probe A or at 43 °C for probe B. After the prehybridization, the filters were hybridized with 5' end-labeled probe A or B for 20 h at 49 or 43 °C, respectively, in the same solution used for the prehybridization. The filters were finally washed in 6  $\times$  SSC (1  $\times$  SSC is 0.15 M NaCl and 0.015 M sodium citrate), 0.1% SDS at 49 or 43 °C for probe A or B, respectively, and exposed to Fuji RX film with intensifying screens at -80 °C. The positive plaques selected on the first screening were further screened with probes A and B under the same conditions used in the first screening.

**Nucleotide Sequencing**—The DNA was extracted from the recombinant phages, digested with *Eco*RI endonuclease, and subcloned into the plasmid vector pUC18. Recombinant plasmids were introduced into *E. coli* MM294 as described by Hanahan (18). Colonies containing the cDNA insert were selected by hybridization with <sup>32</sup>P-labeled probe A. The cDNA inserts of both strands were sequenced in their entirety by the chain-termination method of Sanger *et al.* (19) and by the chemical modification method of Maxam and Gilbert (20).

**RNA Blot Hybridization**—Total RNA was extracted from rice seeds at various stages of maturation as described above. The RNA sample (2  $\mu$ g in total RNA) was denatured and electrophoresed in a formaldehyde-containing agarose gel (21). After electrophoresis, the RNA was transferred to a nitrocellulose membrane (Schleicher & Schuell) and hybridized with nick-translated cDNA (21) at 42 °C in a solution containing 50 mM sodium phosphate (pH 7.0), 5  $\times$  SSC, 50% (v/v) formamide, 0.1% bovine serum albumin, 0.1% polyvinylpyrrolidone, 0.1% Ficoll 400, 0.2% SDS, and 200  $\mu$ g/ml heat-denatured salmon sperm DNA. The filter was finally washed with 0.1  $\times$  SSC containing 0.1% SDS at 42 °C and then exposed.

**Genomic DNA Blot Hybridization**—Rice genomic DNA was isolated from seeds essentially according to the method of Maniatis *et al.* (21). About 6  $\mu$ g of the DNA was digested with various restriction enzymes, fractionated on agarose gels, denatured, and transferred to nitrocellulose membrane according to the method of Southern (22). Filters were prehybridized at 65 °C under the same solution as used for the first screening of cDNA cloning and hybridized at 65 °C in a solution containing denatured nick-translated probes of oryzacystatin or the glutelin (23). The filters were washed with 0.1  $\times$  SSC containing 0.1% SDS at 65 °C.

## RESULTS

**Isolation and Assignment of a cDNA Clone Encoding Oryzacystatin**—From  $1.2 \times 10^5$  independent cDNA clones of a rice seed  $\lambda$ gt10 cDNA library, 12 clones were identified that hybridized with the <sup>32</sup>P-labeled probe A (Fig. 2S). Of these, four clones were also positive to probe B (Fig. 2S) and subjected to further analyses. Since the clone designated as  $\lambda$ OC26 had the longest cDNA insert, its sequence was determined by the strategy shown in Fig. 3S. Fig. 1 shows the complete nucleotide sequence of 598 base pairs which contained the sequences of both probes A and B. It also contained the poly(A) addition signal AATAAA in the 3' region, but

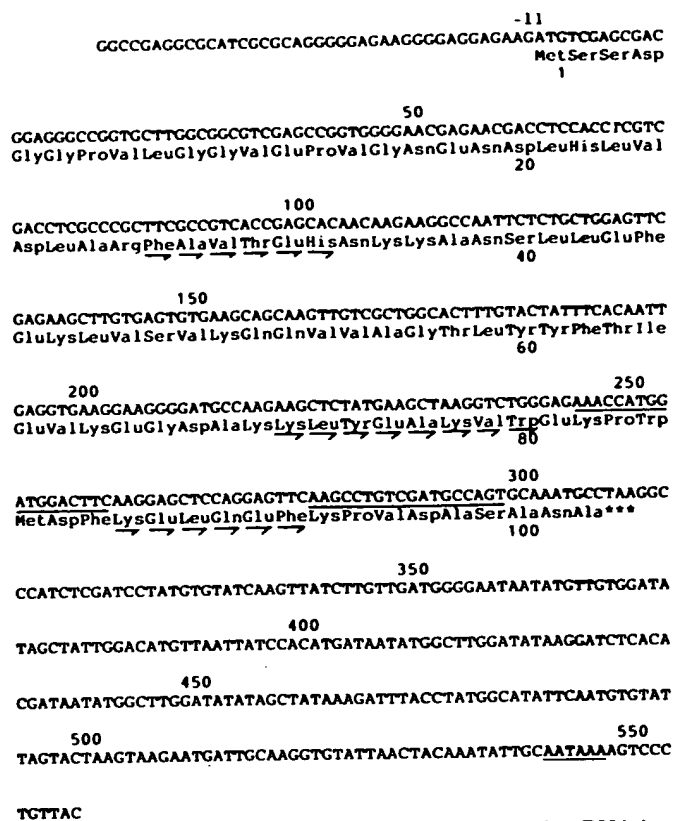


FIG. 1. The nucleotide sequence of oryzacystatin cDNA insert and deduced amino acid sequence. The predicted amino acid sequence is shown below the nucleotide sequence. The nucleotides are numbered in the 5' to 3' direction, beginning with the first nucleotide of the initiation methionine codon ATG; the nucleotides on the 5' side of nucleotide 1 are indicated by negative numbers. The amino acid sequences of two peptide fragments for the synthesis of probes are underlined and other isolated peptides fragments are indicated by arrows. The putative poly(A) addition signal sequence is underlined.

lacked a poly(A) tail. Since one of the six possible reading frames of both strands contained all the six partial amino acid sequences of oryzacystatin (Table 1S), we concluded that clone  $\lambda$ OC26 encoded oryzacystatin in the reading frame shown in Fig. 1. The putative initiation codon (ATG) must be the real initiation codon for the following reasons. 1) The length of the mRNA (about 700 bases) estimated as in Fig. 3 roughly corresponded to the total length of the cDNA insert of  $\lambda$ OC26 and poly(A) tail. 2) Both the molecular weight (11,341) and the amino acid composition (Table I) calculated from the deduced amino acid sequence are consistent with experimental values ( $M_r$  12,000 from SDS-polyacrylamide gel electrophoresis and Table I) (13).

**Amino Acid Sequence Homology**—Comparisons of the deduced oryzacystatin sequence with all proteins in the NBRF database (released Nov. 1986) showed significant sequence homologies with proteins of three cystatin families (Fig. 2). For examples, in the region of residues 8–72 (oryzacystatin numbering), oryzacystatin had a homology of 30.3% with human cystatin A. Residues 8–66 of oryzacystatin was 30.6% homologous with human cystatin C. Residues 26–81 of oryzacystatin also showed a 23.2% homology with human kininogen segment (residues 130–251). Incidentally, it is noteworthy that the conserved sequence Gln-Val-Val-Ala-Gly among several of the animal cystatins exists in oryzacystatin (Fig. 2).

**Expression of Oryzacystatin mRNA during the Ripening Process of Rice Seed**—The cDNA insert of  $\lambda$ OC26 was used as a hybridization probe to evaluate the expression of oryza-

<sup>2</sup> The abbreviation used is: SDS, sodium dodecyl sulfate.

**Identification of the Genes for Oryzacystatin.**—To identify the genes for oryzacystatin and the glutelin, chromosomal DNA of rice seeds was digested with various restriction enzymes and subjected to a total genomic Southern experiment using cDNA probes of oryzacystatin ( $\lambda$ OC26) and the glutelin ( $\lambda$ OG12). Fig. 4 shows profiles obtained in both cases. *Eco*RI and *Pst*I digestion for oryzacystatin (lanes 2 and 5) and *Eco*RI and *Hind*III digestion for the glutelin (lanes 8 and 10) showed essentially single bands, although in both cases some minor bands were detected as well.

## DISCUSSION

<sup>a</sup> The residue numbers and molar percent values are calculated from the sequence deduced (Fig. 1); the initial methionine is included.

The present study also provides useful information concerning phytophysiological significance of oryzacystatin. As the Northern hybridization experiment indicates (Fig. 3), the oryzacystatin mRNA in rice seeds showed its maximal expression at 2 weeks after the flowering of the rice plant and then gradually decreased to an undetectable level at 10 weeks

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50

Oryzacystatin MS SD G G P V L G G V E P V G N E N D L K V D L A R F A V T E H N K K A N S L L - - E F E K L V S V R I Q Q V V A G I

Family 1 cystatins

(a) M I P G G L S E A K P A T P E - - - - - I Q E I V S K V K P Q L E E K T I N E T Y - - G K L E A V Q V K T I Q V V A G I

(b) A C M M C G A P S A T Q P A T A E - - - - - T Q H I A Q Q V R S Q L E E K V N K K F - - P V F K A V S I F K S Q V V A G I

(c) A C M M C G A P S A T H P A T T E - - - - - T Q E I A S K V K S O L E E K A N Q K F - - D V F K A I S F E R R Q V V A G I

Family 2 cystatins

(d) S S P G K P P R L V G G P P D A S V E E G - - - - - V R R A L D F A V G E Y N K A S N D M Y H S R A L Q V V I R A R K Q I V A G I

(e) R L L G G L M E A D V N E E G - - - - - V Q E A L S F A V S E F N K R S N D A Y O S R V R V V I R A R K Q V V I S G I

(f) S E D R S R L L G A P V P V D E N D E G - - - - - L Q R A L G F A M A E Y N R A S N D K Y S S R V R V I S A K R G L V S G I

cystinogen (human)

(g) G P V Y T A Q D Y C L G C V H P I S T Q S P D - - - - - L E P I L R H G I Q Y F N M N T O H S S L F M L N E V K R A G R Q V V A G I

(h) D F V Q P P T K I C V G C F R D I P T N S P E - - - - - L E E I L T H T I T K L N A E M N A I F Y F K I D N V K K A R V Q V V A G I

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50

Oryzacystatin T L Y Y F T I E V H E G D A K L Y E A K - - - - - V Y E K P W M F D K E L E E F K P V A I S A N A

Family 1 cystatins

(a) T N Y Y I K V R A S D N K I M H L A - - - - - V F K S L P G Q N E - - D L V L I G Y O V D K N A D D E L T S F I

(b) T N Y F I K V H V G S E D F V H L R - - - - - V F Q S L P H E N A - - P L I L S Y G I N K A K H D E L T S F I

(c) T N F F I K V D V G E E K C V H L R - - - - - V F E P L P H E N A - - P L T L S S Y O T D K E K H D E L T S F I

Family 2 cystatins

(d) V N Y F L D V I E L S T T I C I A T P N L - - D N C P F H D P H L K R K A F C S F Q I Y A V P S Q G T M L L K S T C C A I

(e) M N Y F L D V I E L S T T I C I S Q A N L - - D S C P F H N P H L K R E L C S F Q I Y V V P W M I N I L V S I F S C G I

(f) I K Y I L G V E I S T T T C P K S G D L - - G S C F H D E P E M A K Y T T C T F V V I S I P W I L N Q I L L E S K G I

cystinogen (human)

(g) L N F R I T Y S I S T N C S K E N F L F L T P D C K - - L A N G E T G E C T D H A Y I D I Q L P I A S F S N - C C I Y P G A

(h) K E F F I C F A P E T T C S E E H E L T E C C E - - T A L S G - - L D C N A E V V V P W E A I Y F T A N - C C I Y P G A

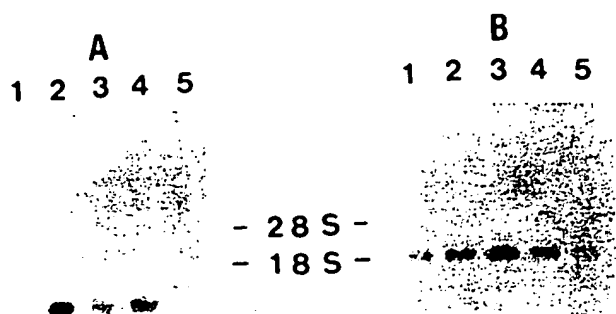


FIG. 3. RNA blot analysis. Two  $\mu$ g each of the total RNA from rice seeds harvested at 1 week (lane 1), 2 weeks (lane 2), 3 weeks (lane 3), 4 weeks (lane 4), or 10 weeks (lane 5) after flowering were electrophoresed, blotted onto a nitrocellulose membrane, and hybridized with  $^{32}$ P-labeled  $\lambda$ OC26 insert (A) or  $^{32}$ P-labeled  $\lambda$ OG10 insert (B). The positions of 18 S and 28 S ribosomal RNA markers are shown ( $\lambda$ OC26 and  $\lambda$ OG10 inserts are probes for oryzacystatin and glutelin, respectively).

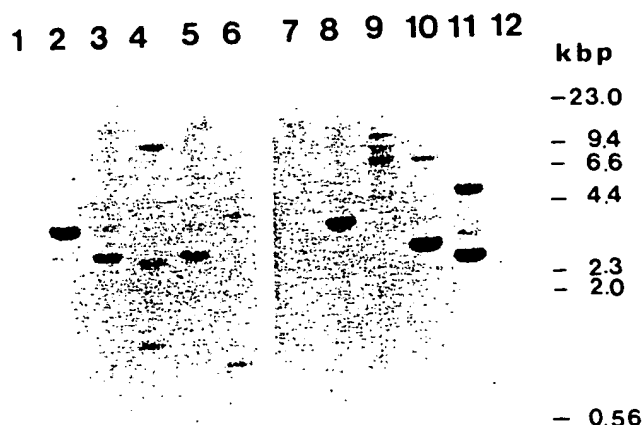


FIG. 4. Southern blot analysis of rice seed genomic DNA. Total rice DNA ( $\sim 6 \mu$ g) was cleaved with *Eco*RI (lanes 2 and 8), *Bam*HI (lanes 3 and 9), *Hind*III (lanes 4 and 10), *Pst*I (lanes 5 and 11), and *Sac*I (lanes 6 and 12), and subjected to blot hybridization using  $\lambda$ OC26 (left) and  $\lambda$ OG12 (right) as probes.

at which time the ripening of the rice seeds is completed. On the other hand, the amount of mRNA encoding rice glutelin, a major storage protein, reaches a maximum at 3 weeks after flowering and then decreases gradually. Thus, the two mRNA species apparently show coordinate expression, although the oryzacystatin mRNA is synthesized and degraded more drastically almost 1 week ahead of the glutelin mRNA. This phenomenon may reflect the existence of a system of regulating the gene expression in rice seeds, involved in controlling the enzymatic proteolysis of the glutelin synthesized during the time from flowering to ripening. We have actually found in rice seeds a cysteine proteinase that efficiently hydrolyzes

glutelin and is almost completely inhibited by an equimolar concentration of oryzacystatin *in vitro* (24).

Knowledge about the structure and mode of expression of this rice cysteine proteinase, as well as those of oryzacystatin, will define a major system involving the proteolytic regulation of rice storage proteins. Gene structural analyses will elucidate the mechanisms for coordinate expression of these proteins.

**Acknowledgments**—We wish to thank Professor Akio Matsuzaki, Experimental Farm in the Faculty of Agriculture at the University of Tokyo, who provided us with the well-defined cultivar of rice, Dr. Hiromichi Nagasawa, Department of Agricultural Chemistry at the University of Tokyo, who cooperated with us in the peptide sequence analysis, and Dr. Hiroshi Kawasaki, Department of Molecular Biology at the Tokyo Metropolitan Institute of Medical Science, who synthesized oligonucleotide probes.

## REFERENCES

- Barrett, A. J., Rawlings, N. D., Davies, M. E., Macheleidt, W., Salvesen, G., and Turk, V. (1986) in *Proteinase Inhibitors* (Barrett, A. J., and Salvesen, G., eds) Vol. 12, pp. 515-569, Elsevier Scientific Publishing Co., Amsterdam
- Macheleidt, W., Borchart, U., Fritz, H., Brzin, J., Ritonja, A., and Turk, V. (1983) *Hoppe-Seyler's Z. Physiol. Chem.* **364**, 1481-1486
- Ritonja, A., Macheleidt, W., and Barrett, A. J. (1985) *Biochem. Biophys. Res. Commun.* **131**, 1187-1192
- Takio, K., Kominami, E., Bando, Y., Katunuma, N., and Titani, K. (1984) *Biochem. Biophys. Res. Commun.* **121**, 149-154
- Takio, K., Kominami, E., Wakamatsu, N., Katunuma, N., and Titani, K. (1983) *Biochem. Biophys. Res. Commun.* **115**, 902-908
- Grubb, A., and Lofberg, H. (1982) *Proc. Natl. Acad. Sci. U. S. A.* **79**, 3024-3027
- Isemura, S., Saitoh, E., Sanada, K., Isemura, M., and Ito, S. (1984) *J. Biochem.* **96**, 1311-1314
- Hirado, M., Tsunasawa, S., Sakiyama, F., Niinobe, M., and Fujii, S. (1985) *FEBS Lett.* **186**, 41-45
- Turk, V., Brzin, J., Longer, M., Ritonja, A., Eropkin, M., Borchart, U., and Macheleidt, W. (1983) *Hoppe Seyler's Z. Physiol. Chem.* **364**, 1487-1496
- Ohkubo, I., Kurachi, K., Takasawa, T., Shiokawa, H., and Sasaki, M. (1984) *Biochemistry* **23**, 5691-5697
- Esterl, W. M., Fritz, H., Kellermann, J., Lottspeich, F., Macheleidt, W., and Turk, V. (1986) in *Cysteine Proteinases and Their Inhibitors* (Turk, V., ed.) pp. 369-392, Walter de Gruyter, Berlin
- Abe, K., Kondo, H., and Arai, S. (1985) *Agric. Biol. Chem.* **49**, 3349-3350
- Abe, K., Kondo, H., and Arai, S. (1987) *Agric. Biol. Chem.* **51**, 2763-2768
- Lewis, W. G., Basford, J. M., and Walton, P. L. (1978) *Biochim. Biophys. Acta* **522**, 551-560
- Brawerman, G., Mandecki, J., and Lee, S. Y. (1972) *Biochemistry* **11**, 637-641
- Aviv, H., and Leder, P. (1972) *Proc. Natl. Acad. Sci. U. S. A.* **69**, 1408-1418
- Gubler, U., and Hoffman, B. J. (1983) *Gene (Amst.)* **25**, 263-269
- Hanahan, D. (1983) *J. Mol. Biol.* **166**, 557-580
- Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H., and Roe, B. (1980) *J. Mol. Biol.* **143**, 161-178
- Maxam, A. M., and Gilbert, W. (1980) *Methods Enzymol.* **65**, 499-560
- Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503-517
- Takaiwa, F., Kikuchi, S., and Oono, K. (1986) *FEBS Lett.* **206**, 33-35
- Abe, K., Kondo, H., and Arai, S. (1987) *Agric. Biol. Chem.* **51**, 1509-1514

by Abe, K., Enori, Y., Kondo, K., Suzuki, K., and Arai, S.

## EXPERIMENTAL PROCEDURES AND RESULTS

**Separation and analysis of peptides from an oryzacystatin digest.** Oryzacystatin was digested with metalloendopeptidase in 50 mM N-methylmorpholine buffer, pH 8.5, at an enzyme/substrate ratio of 1 : 100 (wt/wt) for 18 hr at 37°C. The peptides were fractionated on a Vydac ODS column (Senohu Kagaku) by a linear gradient of 0 to 60% acetonitrile in 1.5 M trifluoroacetic acid at a flow rate of 0.5 ml/min. Fig. 1 shows the results. The amino acid sequences of the numbered peaks were determined as shown in Table 18.

Sequences of the synthetic oligodeoxynucleotide probes used for the selection of oryzaovstatin cDNA clones. The amino acid sequences of peptides corresponding to the nucleotide sequence of the mRNA and the sequences of the two probes (A and B) are shown in Fig. 2-S. The peptides for probes A and B correspond to peaks 6 and 1, respectively (Fig. 1-S).

Restriction map and sequencing strategy for oryocystatin cDNA inserts. The restriction map shows only the relevant sites for DNA sequencing. The protein-coding region is indicated by a filled box in Fig. 3-8. The direction and the extent of sequencing are also shown by horizontal arrows. Sequencing was done by the dideoxy method of Sanger (solid lines) and the chemical modification method of Maxam-Gilbert (dashed lines).

The nucleotide sequence of the rice glutelin cDNA and the predicted amino acid sequence. To compare the 30-mer oligonucleotide corresponding to the glutelin cDNA sequence (23) at nucleotides 11-40 was synthesized. Using this probe, the same cDNA library of immature rice seeds was screened under conditions similar to those described in EXPERIMENTAL PROCEDURES. Several positive clones were isolated and the cDNA insert of one of them ( $\lambda$ OG12) was subcloned into pUC18. Fig. 4-5 shows the result. The partial nucleotide sequence of  $\lambda$ OG12 determined from both the cDNA and the cDNA sequencing method was completely identical to the published nucleotide sequence of glutelin (nucleotides 11-235 and 953-1117) (23), indicating that  $\lambda$ OG12 is a cDNA clone for glutelin.

**Table 18.**  
**Amino acid sequence analyses of fractionated peptides.**

Peak	Amino acid sequence
1	94 102 Lys-Pro-Val-Asp-Ala-Ser-Ala-Asn-Ala 143 137 151 84 160 13 162 121 18
2	73 77 Lys-Leu-Tyr-Glu-Ala 183 168 81 172 165
3	78 80 Lys-Val-Trp- 271 258 71
4	29 34 Xaa-Phe-Ala-Val-Trp-Glu-His- 99 103 80 4 33 13
5	68 73 Lys-Glu-Leu-Gln-Glu-Phe 156 127 105 81 121 48
6	82 87 Lys-Pro-Trp-Met-Asp-Phe 160 93 65 95 25 50

The amounts of PTH-amino acids detected (pmole) are shown below the sequence. The residue numbers of amino acids are shown above the sequence.

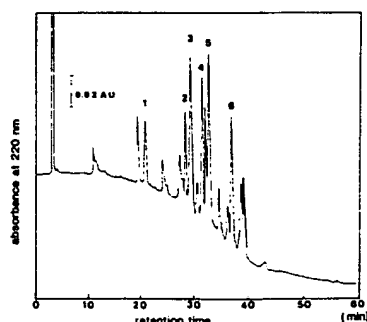


Fig. 1-S. Separation of peptides from an oryzacystatin digest.

		62				67
Peptide sequence :		Lys	Pro	Trp	Met	Asp
mRNA sequence :	5'	AAA	CCU	UUG	AUG	GAU
		G	C			C
			A			
			G			
Probe A	3'	TTT	GGA	TCC	TAC	AA
(17mer, 16mix)		C	G		G	
			T			

Peptide sequence :	94					99	
	Lys	Pro	Val	Asp	Ala	Ser	
mRNA sequence :	5'	AAA	CCU	GUU	GAU	GCU	UC 3'
		G		C	C	C	AG
			A	A			
			G	G		G	
Probe B	3'	TTT	GGA	CAA	CTA	CGA	AG 5'
(17mer, 1024ms)		C		G	G	G	TC
			T	T			
			C	C		C	

Fig. 2-5. The sequences of the synthetic oligodeoxynucleotide probes used for the selection of oryxacystatin cDNA clones.

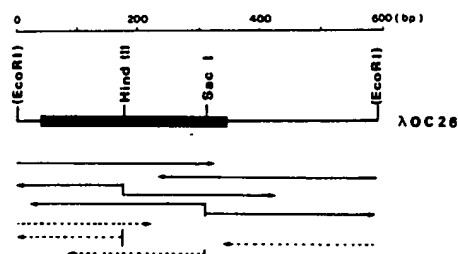


Fig. 3-S. Restriction map and sequencing strategy for oryzacystatin cDNA inserts.

[illegible]

Fig. 4-S. The nucleotide sequence of the rice glutelin cDNA and the predicted amino acid sequence.

Claim 63

$\alpha$ -amylase inhibitor

# Selective expression of a probable amylase/protease inhibitor in barley aleurone cells: Comparison to the barley amylase/subtilisin inhibitor

J. Mundy<sup>1</sup> and J.C. Rogers<sup>2</sup>

<sup>1</sup> Department of Biotechnology, Carlsberg Research Laboratory, Gamle Carlsberg Vej 10, DK-2500 Valby, Copenhagen, Denmark, and

<sup>2</sup> Division of Hematology/Oncology, Departments of Internal Medicine and Biology, Washington University School of Medicine, St. Louis, MO 63110, USA

**Abstract.** We have cloned and sequenced a 650-nucleotide cDNA from barley (*Hordeum vulgare* L.) aleurone layers encoding a protein that is closely related to a known  $\alpha$ -amylase inhibitor from Indian finger millet (*Eleusine coracana* Gaertn.), and that has homologies to certain plant trypsin inhibitors. mRNA for this probable amylase/protease inhibitor (PAPI) is expressed primarily in aleurone tissue during late development of the grain, as compared to that for the amylase/subtilisin inhibitor, which is expressed in endosperm during the peak of storage-protein synthesis. PAPI mRNA is present at high levels in aleurone tissue of desiccated, mature grain, and in incubated aleurone layers prepared from rehydrated mature seeds. Its expression in those layers is not affected by either abscisic acid or gibberellic acid, hormones that, respectively, increase and decrease the abundance of mRNA for the amylase/subtilisin inhibitor. PAPI mRNA is almost as abundant in gibberellic acid-treated aleurone layers as that for  $\alpha$ -amylase, and PAPI protein is synthesized in that tissue at levels that are comparable to  $\alpha$ -amylase. PAPI protein is secreted from aleurone layers into the incubation medium.

**Key words:** Absciscic acid and mRNA induction – Aleurone – Amylase/Protease inhibitor – Gibberellin and mRNA induction – *Hordeum* (aleurone protein).

**Abbreviations:** ABA=abscisic acid; ASI=barley amylase/subtilisin inhibitor; bp=nucleotide base pairs; Da=dalton; dpa=days post anthesis; GA<sub>3</sub>=gibberellic acid; PAPI=probable amylase/protease inhibitor; poly(A)RNA=polyadenylated RNA; SDS-PAGE=sodium dodecyl sulfate-polyacrylamide gel electrophoresis

## Introduction

Our laboratories are interested in the mechanisms by which specific genes are expressed under developmental and hormonal controls in starchy endosperm and in aleurone tissue of barley (*Hordeum vulgare* L.). An approach to understanding this problem can be initiated by identifying and characterizing the genes for proteins that are expressed in abundance in specific tissues at specific developmental times. In this regard, much is known about the synthesis in developing endosperm of prolamin storage polypeptides (Mathews and Mifflin 1980; Giese and Hopp 1984) and of several salt-soluble proteins (Jonassen et al. 1981; Ponz et al. 1983; Giese and Hejgaard 1984), but relatively little is known about the pattern of gene expression and of protein synthesis in developing aleurone tissue. Being derived developmentally from the endosperm, any separate function of the aleurone during grain development is not clear, and there are no data on a molecular level that clearly identify processes separate from those in progress in the endosperm. After rehydration of the desiccated grain however, the aleurone layer assumes a new role; its function is to destroy the now dead tissue from which it derived. In response to gibberellins produced by the embryo during germination, the aleurone layer synthesizes and secretes large quantities of hydrolases which mobilize the endosperm reserves for use by the growing plant. These hormonally regulated events have been studied extensively (see Jacobsen 1984 for a recent review).

There has been no evidence that aleurone tissue produces hydrolase-inhibitory proteins in any abundance. In contrast, many of the abundant

proteins in the endosperm of barley and other cereals are inhibitors of  $\alpha$ -amylases and of proteases (Shivaraj and Pattabiraman 1981; Boisen et al. 1981; Odani et al. 1983; Weselake et al. 1983; Mundy et al. 1983). Recent work has shown that some of these proteins are closely related (Odani et al. 1983; Campos and Richardson 1983; Shewry et al. 1984). The inhibitors of trypsin from barley and of  $\alpha$ -amylases from wheat share homologies with a bifunctional inhibitor of both  $\alpha$ -amylase and trypsin from finger millet (*Eleusine coracana* Gaertn.) (Campos and Richardson 1983). All of them may now be considered members of the barley trypsin-inhibitor family (Odani et al. 1983). Other bifunctional inhibitors have been found in barley and wheat which inhibit endogenous cereal  $\alpha$ -amylase 2 isoenzymes as well as bacterial subtilisins (Mundy et al. 1983; Weselake et al. 1984). The amino-acid sequences of these proteins, called ASI for amylase/subtilisin inhibitors, are homologous to the sequence of the soybean trypsin inhibitor (Kunitz) and are therefore classified as members of that family (Mundy et al. 1984; Hejgaard et al. 1983).

The precise role of these inhibitors is not known, but the general assumption is that they may protect the grain from exogenous hydrolases. The ASIs, as inhibitors of the seed's own "germination-specific"  $\alpha$ -amylase (Weselake et al. 1983; Mundy et al. 1983), may also help to regulate the activity of this enzyme which is important in mobilizing the seed's starch reserves during germination. In addition, protease inhibitors may contribute towards the establishment of a balanced amino-acid pool in the endosperm storage proteins (Jonassen 1980; Hejgaard et al. 1985).

In this paper, we describe a novel amylase/protease inhibitor-type gene, identified by both its mRNA and protein, that is the first shown to be expressed primarily in aleurone tissue late during grain development, and that is expressed there at high levels after rehydration of the desiccated seed, regardless of the presence or absence of gibberellic acid ( $GA_3$ ) or abscisic acid (ABA). The specificity of expression of this gene is established by comparing it to the expression of barley ASI, which is shown here to be primarily an endosperm protein and which is induced by ABA in mature aleurone layers incubated in short-term tissue culture.

## Material and methods

**Tissue preparation.** Barley (*Hordeum vulgare* L., cv. Himalaya, 1979 crop) seeds were obtained from the Department of Agronomy, Washington State University, Pullman, USA. Mature

aleurone layers were prepared from imbibed grains and incubated as described in Rogers and Milliman (1983); for incubations with hormones, the buffer contained either  $1 \cdot 10^{-6}$  M  $GA_3$  or  $10^{-5}$  M ABA (both hormones from Sigma Chemical Co, St. Louis, Mo., USA).

**Separation of aleurone and starchy endosperm tissues.** Developing grains were harvested at 20 and 30 days post anthesis (dpa), which correspond to the peak and end of storage-protein synthesis, respectively (Mathews and Mifflin 1980; Giese and Hejgaard 1984). At these stages, the developing starchy endosperms can be easily squeezed free of the surrounding aleurone-pericarp tissue after removal of the embryo. Light and scanning electron microscopy (Mundy et al. 1986) showed that these endosperms were free of aleurone tissue. The aleurone layers, however, often have several layers of endosperm cells attached to them. These cells were removed by longitudinally cutting the layers open and scrapping them with a scalpel. Following three washes in, phosphate-buffered saline (PBS; 0.1 M sodium phosphate, pH 7.2, 0.015 NaCl), they were essentially free of endosperm cells.

It was not possible to obtain aleurone and endosperm tissues by dissection in the dry seed because they are too tightly bound to each other. We therefore used mechanical abrasion to remove and collect the outer layers of the seed. Mature starchy endosperm and aleurone tissues were prepared by progressively removing the outer portions of mature, dry grain in a rice polisher equipped with a file disc (Pearlest; Kett Electric Laboratory, Tokyo, Japan). The aleurone fraction represented the most external 3–20% fractional volume (inclusive) of the grains removed by abrasion, while the endosperm fraction represented the innermost 30–100% fractional volume (inclusive). The extent of aleurone cell removal was assessed by examining the pearled seeds in an epifluorescence microscope. This sensitive, qualitative method takes advantage of the fact that aleurone cell walls are specifically rich in ferulic acid which fluoresces strongly in ultraviolet light (Jensen et al. 1982). Endosperms pearled to 30% were free of aleurone cells except in recessed parts of the crease. These areas were removed with a scalpel.

**Isolation of and cDNA cloning RNA.** Methods for isolation of RNA from aleurone layers, cDNA cloning, RNA and DNA gel electrophoresis, and DNA sequence analysis were as described in Rogers (1985). A complementary-DNA library containing 8000 clones were prepared from RNA isolated from aleurone layers that had been incubated for 20 h in the presence of ABA; it was screened for clones carrying sequences abundant in that tissue by colony hybridization with  $^{32}P$ -labeled cDNA prepared from the same RNA source. RNA was isolated from other tissues utilizing the same method as for aleurone layers.

**Antibody preparation.**  $\alpha$ -Amylase and ASI were purified from malt (Mundy et al. 1983) and PAPI was purified from barley grains according to Svensson et al. (1986). Antibodies towards  $\alpha$ -amylase and ASI were raised according to standard procedures, while antibody towards PAPI was prepared by immunization after glutaraldehyde polymerization of the antigen (Bollum 1975). Immunoglobulins were purified from the crude antisera by chromatography on Protein-A Sepharose 4B (Pharmacia Corporation, Stockholm, Sweden) by standard methods, and the protein concentration of the antibody preparations adjusted to  $4 \text{ mg} \cdot \text{ml}^{-1}$ .

**Extraction and immune-blotting of proteins.** Lyophilized tissues were ground in a mortar and pestle with 10 vol. of water, extracted by rotating slowly for 3 h at room temperature, and then centrifuged at  $10000 \cdot g$  for 5 min. Polypeptides in extract

aliqu  
by s  
(SDS  
3-an  
pH  
Mill  
incu  
pero  
(Tov

Anal  
deny  
tant  
SDS  
nine  
(Arli

with  
prior  
son a  
tratic  
600  
1·10<sup>6</sup>  
ratio  
aleur  
freez  
the s  
in th  
extra  
(4.5  
ylsul  
acid  
buffe

Meth  
ical  
in th  
puter  
with  
amyl  
was n  
tein s  
made  
PAM

Resi

Isol  
lase,  
ry tr  
filter  
thesi  
untr  
a m  
Mill  
fied  
catin  
abu  
rone  
clon  
iden  
The  
arbi



aliquots containing equal amounts of protein were separated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), and transferred electrophoretically in 0.15 M 3-amino-2-(hydroxymethyl)-1,3-propanediol-(Tris)-glycine, pH 8.9, 20% methanol to nitrocellulose paper (type HAHY; Millipore Corp., Bedford, Mass., USA). The paper was then incubated with  $10 \mu\text{g} \cdot \text{ml}^{-1}$  rabbit antibody and assayed with peroxidase-labelled swine anti-rabbit immunoglobulins (IgG) (Towbin et al. 1979).

**Analysis of polypeptides synthesized in vivo and in vitro.** Polyadenylated RNA (poly(A)RNA) was translated, and the resultant protein products were analyzed by immunoadsorption and SDS-PAGE as described in Mundy et al. (1985) [ $^{35}\text{S}$ ]Methionine and rabbit reticulocyte lysate (N 90) were from Amersham (Arlington Heights, Ill., USA).

Proteins in incubated aleurone layers were labelled in vivo with [ $^{35}\text{S}$ ]methionine and extracted into buffers containing SDS prior to dilution and immunoadsorption as described in Anderson and Blobel (1983) and Mundy et al. (1985). Extract concentrations of  $0.25 \cdot 10^6$ ,  $1 \cdot 10^6$ ,  $5 \cdot 10^6$ ,  $10 \cdot 10^6$  and  $20 \cdot 10^6$  cpm per 600  $\mu\text{l}$  immunoadsorption reaction mixture (specific activity  $1 \cdot 10^4$ – $1.2 \cdot 10^4$  cpm  $\cdot \mu\text{g}^{-1}$  protein) were used to optimize the ratio of antigen to antibody. To measure proteins secreted by aleurone cells, incubation media were first concentrated by freeze-drying, then dissolved in SDS buffer, and processed in the same way as the extracts. Because the amount of protein in the incubation media was much lower than that in the tissue extracts, only media aliquots containing  $1 \cdot 10^6$  cpm ( $4.5 \cdot 10^4$ – $13.7 \cdot 10^4$  cpm  $\cdot \mu\text{g}^{-1}$  protein) were used. Phenylmethylsulfonyl fluoride (100  $\mu\text{M}$ ) and ethylenediaminetetraacetic acid (EDTA; 5 mM) were included in the immunoadsorption buffers to minimize proteolytic degradation.

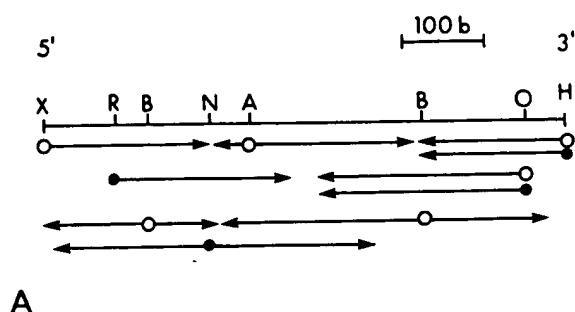
**Methods for sequence-homology analysis.** The National Biomedical Research Foundation Protein Sequence Databank carried in the Washington University Medical School VAX-11 computer was screened for sequences related to the PAPI protein with the SEARCH program of Dayhoff et al. (1983). The Ragi amylase inhibitor 1–2 sequence (Campos and Richardson 1984) was not present in the Databank. Detailed comparisons of protein sequences thought to be related to the PAPI sequence were made with ALIGN, utilizing the mutation data matrix (250 PAMs) +6, and a gap penalty of 8 (Dayhoff et al. 1983).

## Results

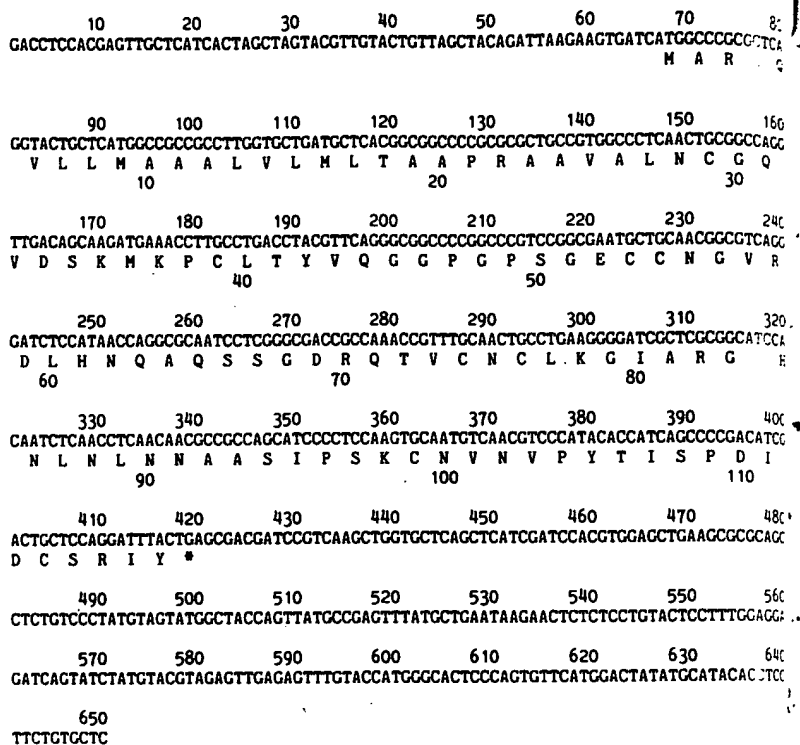
**Isolation of a cDNA representing a probable amylase/protease inhibitor.** Replica plates of the primary transformants were transferred to nitrocellulose filters and hybridized with  $^{32}\text{P}$ -labeled cDNA synthesized from poly(A)RNA isolated from either untreated or from ABA-treated aleurone layers in a manner similar to that described in Rogers and Milliman (1983). Seven separate clones were identified that hybridized strongly to both probes, indicating that they represented a mRNA that was abundant in either untreated or ABA-treated aleurone tissues. Plasmids isolated from all seven clones contained cDNA inserts with apparently identical sizes and preliminary restriction sites. Therefore one of the seven, identified as A1f, was arbitrarily selected for more detailed characteriza-

tion. The cDNA insert was sequenced in its entirety by the chemical degradation method (Maxam and Gilbert 1980); both strands were sequenced and all restriction-enzyme sites used for end labelling were confirmed by sequencing (Fig. 1A). The complete nucleotide sequence, and the predicted sequence of the protein product are presented in Fig. 1B. Because of the homology of the predicted protein sequence to known amylase and protease inhibitors (see below), it is referred to subsequently as PAPI, for probable amylase/protease inhibitor. The cDNA insert contains 650 base pairs (bp), not including the poly(dA) tail. The first ATG is preceded by a 67-bp leader sequence, and is followed by an open reading frame that terminates with TGA at nucleotide 419. The coding region is 64% G + C, a feature shared by both  $\alpha$ -amylase cDNAs (Rogers and Milliman 1983; Rogers 1985) and the thiol protease cDNA (Rogers et al. 1985). At the 3'-end, this sequence lacks the AATAAA polyadenylation signal usually found a short distance preceding the poly(dA) tail, although the sequence AATAAG is present beginning at position 528, 122 nucleotides preceding the poly(dA) sequence. This sequence has been reported in a legumin gene near the stop codon in a mRNA that also had the canonical AATAAA near the poly(A) tail (Croy et al. 1982), but the arrangement seen here appears to be one not previously reported. Although this cDNA has a long 5' untranslated region, we do not know that is, in fact, represents the full-length sequence of the mRNA, since primer extension and  $\text{S}_1$  nuclease analyses have not been performed. Southern blots carrying barley genomic DNA digested with restriction enzymes not cutting within the cDNA sequence (such as *Hind*III and *Eco*R1) were probed with the labelled cDNA; single hybridizing bands were obtained, indicating that probably only one gene is present (data not presented). Thus we proceeded under the premise that our experiments measured the expression of at least a very small gene family, and most likely a single gene. Detailed analysis of multiple genomic clones will be needed to resolve the question of the exact number of homologous sequences.

Svensson et al. (1986) purified a 10-kDa basic protein from barley grains and determined the sequence of the first 37 N-terminal and the last two C-terminal amino acids. These sequences match that presented in Fig. 1B exactly, beginning with residue 26. As noted by Svensson et al. (1986), the predicted protein sequence is closely related to that for a known  $\alpha$ -amylase inhibitor from Ragi or Indian finger millet (Campos and Richardson 1984);



**Fig. 1. A** Sequencing strategy for the cDNA insert. Restriction-enzyme fragments were end-labeled at the sites shown and sequence determinations were obtained for the regions indicated by the arrows. Restriction sites were: X, Xba I (in pUC18 polylinker); R, Rsa I; B, BssH II; N, Nci I; A, Ava I; O, Nco I; and H, HindIII (in pUC18 polylinker). Fragments were labeled at the 5' (●), or the 3' (○) ends. **B** Nucleotide and derived amino-acid sequence of the cDNA insert. The insert contains 650 nucleotides excluding the poly(dA) tail. The amino-acid sequence is presented for the only open reading frame and begins with the first ATG initiation codon



the alignment of these sequences is presented in Fig. 2A. It can be seen that the two sequences, beginning with PAPI residue 26, have 45 out of 96 residues identical, with two single residue gaps in the PAPI sequence necessary for optimal alignment. The first 25 amino acids, which include a hydrophobic core, probably represent a signal peptide, since they are absent from the mature PAPI protein. In addition, the PAPI sequence is homologous to known trypsin inhibitors. Presented in Fig. 2B are comparisons to two Bowman-Birk type inhibitors, one from eggplant (*Solanum melongena* L.), and one from *Macrotyloma axillare* (family Leguminosae) (Joubert et al. 1979). Bowman-Birk inhibitors are small proteins comprised of two repeating units, each unit having a different inhibitory specificity (Laskowski and Kato 1980). Different regions in the PAPI sequence have homology to regions containing the active-site residues in the trypsin-inhibitory domains of the two proteins. In addition, it can be seen (Fig. 2B, top) that apparently both N-terminal and C-terminal portions of the PAPI sequence are related to the *Macrotyloma* protein's trypsin-inhibitory domain, although in the C-terminal portion, alignment includes a lysine residue corresponding to the active site. This type of protein does not fit with any of the previously established groups of protease inhibitors (Laskowski and Kato 1980).

The availability of this protein, derived from the same gene as our cloned cDNA, and antiserum to it, presented the opportunity to study the expression of this gene at multiple levels. We therefore studied expression of the gene by assessing the relative abundance of hybridizable RNA in different tissues. The presence of active mRNA was assessed by translating mRNA from the same tissues in vitro and immunoprecipitating PAPI. The expression of this mRNA in vivo was assessed by labeling with [<sup>35</sup>S]methionine, and by determining the relative abundance of PAPI in different tissues by immune blotting.

*PAPI RNA is expressed constitutively at high levels in incubated aleurone layers.* PAPI-cDNA labeled by nick translation hybridized strongly to an approx. 700-nucleotide RNA in both untreated (Fig. 3A: U) and ABA-treated (Fig. 3A: A) aleurone tissue cultures. In contrast, RNA from either shoot (Fig. 3A: S), or root (Fig. 3A: R) contained no detectable hybridizing RNA, even on prolonged autoradiographic exposure of the blot (data not presented). The specificity of this hybridization pattern was demonstrated by hybridizing identical blots with probes for  $\alpha$ -amylase (Rogers and Millman 1983) (Fig. 3A: amylase) or for the GA<sub>3</sub>-induced thiol protease (Rogers et al. 1985) (Fig. 3A: protease). Consistent with previous results (Rogers

A. Comparison between the PAPI sequence and the  $\alpha$ -amylase inhibitor from Indian Finger Millet (Ragi):

```

      10      20      30      40      50
PAPI:  MARAQVLLMAAALVLM LTAAPRAAVALNCCQVDSKMKPCLTYVQG-GPGFSGECCNGVRDL
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
RAGI:  AISCCQVSSAIGPCLAYARGAGAAPSASQSGVRS
      10      20      30

      70      80      90      100     110
HNQAQSSGDRQTVNCN-LKGIARGIHNLNNAASIPSKCNVNPVYTISPDI DCSRIY
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
NAAARTTADRRRAACNSLKSASRSVGLNAGKASSIPGRGVR LPYAISASIDCSRVNN
      40      50      60      70      80      90

```

B. Comparison of PAPI to regions from two Bowman-Birk type trypsin inhibitors containing the active site residues:

1. Inhibitor DE-4 from *Macrotyloma axillare*

		K P C		C C N R	SCORE
PAPI	35-K M K P C L T Y V Q G G P G P S G E C - C N G V R-58				3.86 (<.0001)
DE-4	11-S S K P C C D L C T C T K S I P P Q C H C N D M R-35				
PAPI	66-S S G D R Q T V C N C L K G I A R G I H N L N L N-90				3.70 (.0001)
Common	S S C C K I H				

2. Eggplant trypsin inhibitor

		SCORE
Eggplant	29-N P E N P K A C P R N C D G R I A Y G I C P L S-52	
PAPI	88-N L N N A A S I P S K C N V N V P Y T I S P D I-111	5.02 (<.000001)
Common	N N P C Y I P	

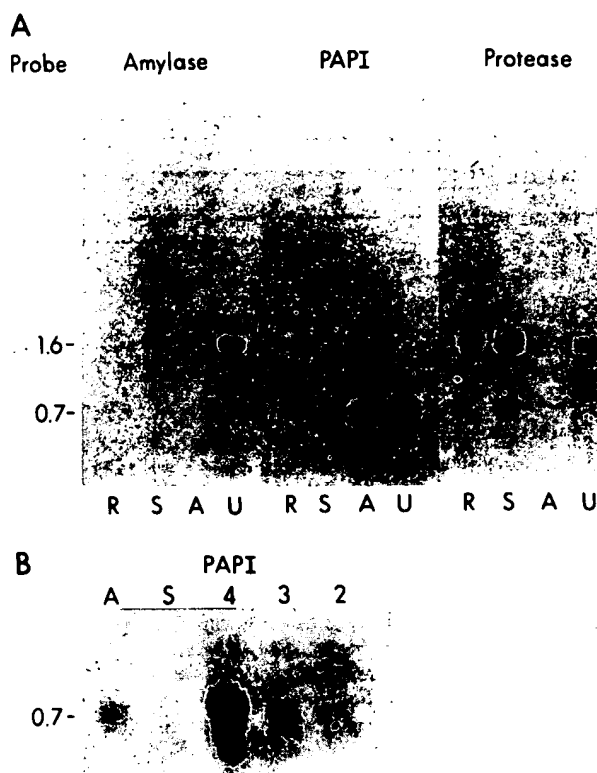
Fig. 2A, B. Comparison of the PAPI amino-acid sequence to sequences for (A) an  $\alpha$ -amylase inhibitor from ragi (Campos and Richardson 1984), and (B) protease inhibitors from *Macrotyloma axillare* (Joubert et al. 1979) and eggplant (Richardson 1979). Computer programs for searching the National Biomedical Foundation Protein Sequence Databank (SEARCH), and for optimal alignment and statistical evaluation of similarity (ALIGN) were from Dayhoff et al. (1983) and used as described in *Material and methods*. The protease-inhibitor sequences were selected for analysis because they were among the most likely candidates for homologous proteins as identified by SEARCH. The numbers bracketing the sequences in B designate the first and last residues, respectively, in the sequence. Asterisks indicate the active-site residues for the known protease inhibitors. Residues shared by the two sequences (common) are presented above or below the comparison. The alignment "Score" represents the number of standard deviations a particular comparison rates above that from comparison of randomly generated sets of sequences with the same amino-acid composition; the probability that a given score would occur on only a chance basis (Dayhoff et al. 1983) is presented in parentheses. In A, the alignment score was >14, with a probability of <10<sup>-12</sup> that it was the result of chance

1985; Rogers et al. 1985), the  $\alpha$ -amylase probe detects mRNA that is present in untreated aleurone cells and is present at very low levels in shoot and root, while the protease probe detects mRNA that is present in untreated aleurone tissue but also is present in more abundance in shoot and root. The abundance of both mRNAs is decreased when the aleurone layers are treated with ABA.

Hybridizable RNA can also be detected with the PAPI probe late during development of aleurone plus endosperm tissue. In Fig. 3B, in comparison to RNA from untreated mature aleurone layers (lane U), little or no hybridizable RNA is present in endosperm-aleurone tissue harvested two weeks post anthesis (Fig. 3B, lane 2), but in-

creasing quantities are detected at three (lane 3) and four (lane 4) weeks post anthesis, such that the relative abundance in the four-week post anthesis sample is greater than that detected in the samples from incubated aleurone layers. At approximately five weeks the grain is desiccated and mature.

In order to estimate the abundance of the PAPI mRNA in incubated aleurone layers, an experiment was performed comparing it to the abundance of  $\alpha$ -amylase mRNA. Although no quantitative hybridization experiments have been performed that would allow an estimate of the number of amylase transcripts per cell, in-vitro translation experiments have demonstrated that these appear



**Fig. 3A, B.** Distribution of hybridizable PAPI RNA in different barley tissues. 10- $\mu$ g samples of total RNA were subjected to electrophoresis through a 1% agarose gel in the presence of formaldehyde, transferred to nitrocellulose, and hybridized (see *Material and methods*). **A** Hybridization to RNA from mature excised, incubated aleurone layers and to RNA from shoot and root tissue. Three identical blots were prepared and hybridized with nick-translated plasmid DNA carrying the  $\alpha$ -amylase clone E (Rogers and Millman 1983) insert (*left*), the PAPI cDNA (*center*), and the aleurain thiol protease cDNA (Rogers et al. 1985) (*right*). The RNA samples were obtained from root (R) or shoot (S) tissue from recently germinated plants, or from excised, mature aleurone layers incubated for 20 h in the absence (U) or presence (A) of 10  $\mu$ M ABA. The numbers to the side indicate the estimated size in kilobases of the hybridizing RNA; these estimates were derived from bacteriophage  $\lambda$  Hind-III/EcoRI fragments subjected to electrophoresis in parallel and hybridized separately on a different blot. **B** Hybridization to RNA from developing aleurone + starchy endosperm. Total RNA was isolated from the deembryonated grain two, three and four weeks post anthesis, and the abundance of hybridizable PAPI RNA was compared to that in ABA-treated aleurone (A) and shoot (S) RNAs as in A.

to be the most abundant mRNAs present (Mozer 1980; Higgins et al. 1982). The comparison here (Fig. 4) utilized hybridization of cDNA probes, made either from poly(A)RNA from untreated or GA<sub>3</sub>-treated aleurone layers, to dot blots of equal quantities of plasmid DNAs carrying the  $\alpha$ -amylase clone E (Fig. 4, lanes E), PAPI (lanes labeled A), and thiol-protease (lanes labeled G) inserts.

The assumption is that all mRNAs present in the reverse-transcriptase reaction mixture will be transcribed with a frequency proportional to their abundance, and therefore the amount of radioactive label incorporated will reflect those proportions. (This assumption is useful only in a semi-quantitative way because there are variables, such as G+C content and secondary structure, that could alter the efficiency of transcription.) In a filter hybridization, where equal amounts of target sequences are present, the hybridization signal detected by direct autoradiography to each of the target sequences will reflect their relative abundances in the probe used. (Another correction should take account of the fact that the amylase cDNA is twice as large as that for PAPI, and thus should have twice the hybridization signal for a given molar amount. This is ignored here because of uncertainty about the average length of the labeled cDNAs used for hybridization.) Thus, the intensity of hybridization to dots containing the PAPI plasmid DNA can be compared to that of dots containing amylase sequences to determine the relative proportions of those two species in the cDNA probe, and, within the constraints of the problem of reverse-transcription efficiency, can be used to estimate their relative abundance in cellular mRNA.

The results of such an experiment are presented in Fig. 4A and B. Equal quantities (by weight) of amylase (E), PAPI (A), and thiol protease (G) plasmid DNAs were spotted on two identical filters and hybridized with cDNA probes from untreated (Fig. 4A), or GA<sub>3</sub>-treated (Fig. 4B) aleurone-cell mRNA. It can be seen that PAPI sequences appear to be the most abundant of the three in untreated cell RNA, while in GA<sub>3</sub>-treated RNA, the abundance of the amylase sequences appears to be at most fourfold greater than those for PAPI. Thus, the PAPI transcripts must be one of the most abundant in these cells. To establish that GA<sub>3</sub> treatment of aleurone tissue cultures did not appreciably change the abundance of PAPI transcripts, a dot blot containing equal quantities of total RNA isolated from untreated (Fig. 4C: U) cells, or cells that had been incubated in the presence of 2.5  $\mu$ M GA<sub>3</sub> for 12 (Fig. 4C: 12) or 24 (Fig. 4C: 24) hours, was hybridized with nick-translated PAPI cDNA. The relative amounts of PAPI sequences detected in those RNAs were similar. These results, in aggregate, indicate that PAPI mRNA is constantly expressed in aleurone layers isolated from rehydrated mature grains, and its abundance is not altered to any significant degree by either of the two plant hormones, GA<sub>3</sub> or ABA.

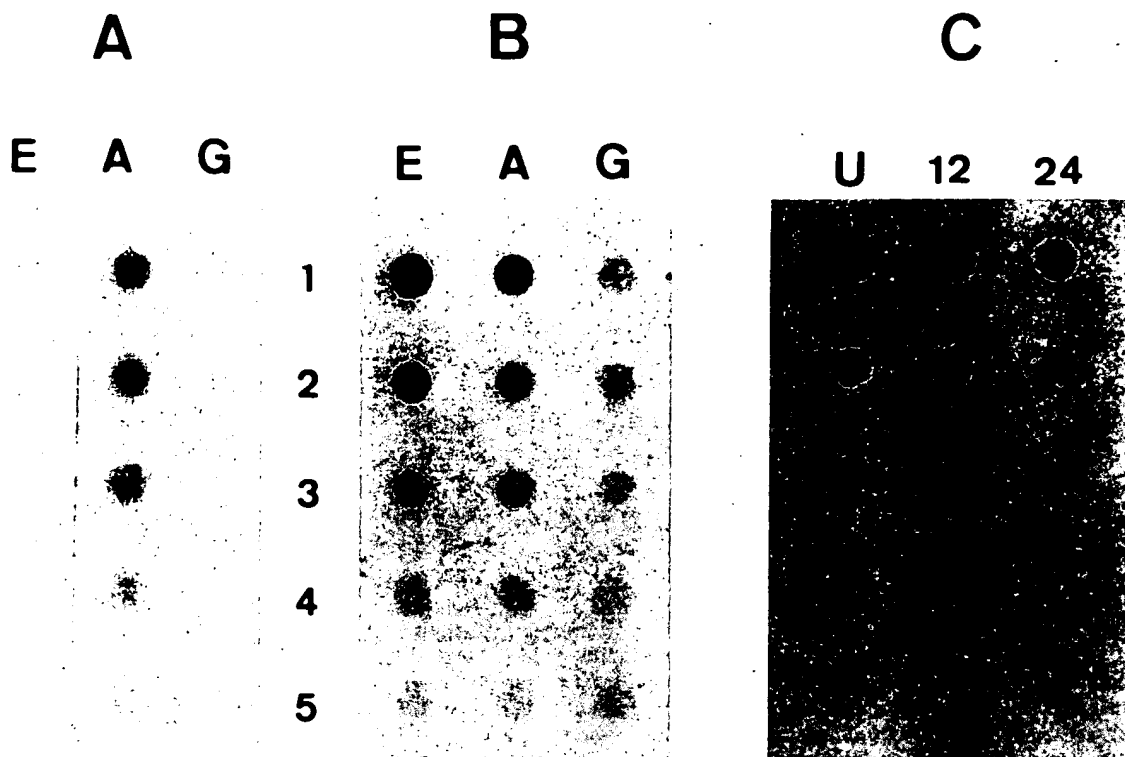


Fig. 4A-C. Relative abundance of PAPI RNA in mature barley aleurone layers after incubation under different conditions. Panels A and B represent dot blots in which plasmid DNA was fixed to nitrocellulose and hybridized with oligo(dT)-primed,  $^{32}\text{P}$ -labeled cDNA synthesized from poly(A)RNA isolated from aleurone layers incubated in the absence (A) or presence of  $2.5\ \mu\text{M}$   $\text{GA}_3$  (B) for 18 h. The plasmid DNAs were: E, clone E  $\alpha$ -amylase cDNA (Rogers and Milliman 1983); A, PAPI cDNA; G, aleurain thiol protease cDNA (Rogers et al. 1985). In each instance, the total amount of DNA applied to each dot was  $1\ \mu\text{g}$ ; the excess over the plasmid DNA amount was made up of carrier salmon-sperm DNA. The amounts of plasmid DNA applied were: line 1,  $1\ \mu\text{g}$ ; 2,  $0.25\ \mu\text{g}$ ; 3,  $0.06\ \mu\text{g}$ ; 4,  $0.013\ \mu\text{g}$ ; and 5,  $0.003\ \mu\text{g}$ . In these experiments the standard hybridization buffer also included  $5\ \mu\text{g}\cdot\text{ml}^{-1}$  of poly(dA). In C total RNA from different sources was fixed to the filter and hybridized with nick-translated PAPI plasmid DNA. The RNA was derived from untreated aleurone layers incubated for 20 h (U), or from layers that were treated with  $\text{GA}_3$  for 12 or 24 h. Each dot contained  $1\ \mu\text{g}$  of RNA, with the difference above that from aleurone RNA made up of carrier yeast tRNA. The amounts of aleurone RNA for each line were; line 1,  $1\ \mu\text{g}$ , and the remainder, fourfold dilutions in amounts identical to those used for plasmid DNAs in A and B

*Translatable PAPI mRNA in seed tissues and excised, mature aleurone layers.* We wanted to compare the relative abundance of PAPI mRNA to that for ASI in seed tissues at various developmental stages. This comparison would allow us to determine if the expression of PAPI differed from that of other abundant proteins synthesized during grain development. This seemed to be an appropriate comparison because ASI is known to be an amylase/protease inhibitor, and is known to be present in both mature grain (Mundy et al. 1983) and excised, mature aleurone layers (Mundy 1984). Since no cDNA clone for ASI was available to us, hybridization experiments were not possible. Therefore the presence of mRNA for the two proteins in different tissues in developing and mature seed was assessed by isolating poly(A)RNA from those tissues, translating it in a reticulocyte system

in the presence of [ $^{35}\text{S}$ ]methionine, and specifically immunoprecipitating the two labeled proteins.

Figure 5 presents autoradiograms of SDS-PAGE analysis of proteins synthesized by translation of mRNA in vitro. It can be seen that mRNA for ASI is readily detected in 20- and 30-dpa starchy endosperm (Fig. 5B, lanes 1 and 2), while little or no PAPI mRNA is present (Fig. 5C, lanes 1 and 2). In contrast, in endosperm from dry seed (lanes 3), no ASI mRNA could be detected, but relatively abundant PAPI mRNA was present. This demonstrates that ASI mRNA is present in endosperm at times corresponding to the peak of storage-protein synthesis, while PAPI mRNA appears much later. In addition, in the developing grain, aleurone tissue shows striking specificity for expression of PAPI. No ASI mRNA was detected in aleurone tissue from 20 dpa (Fig. 5B, lane 4),

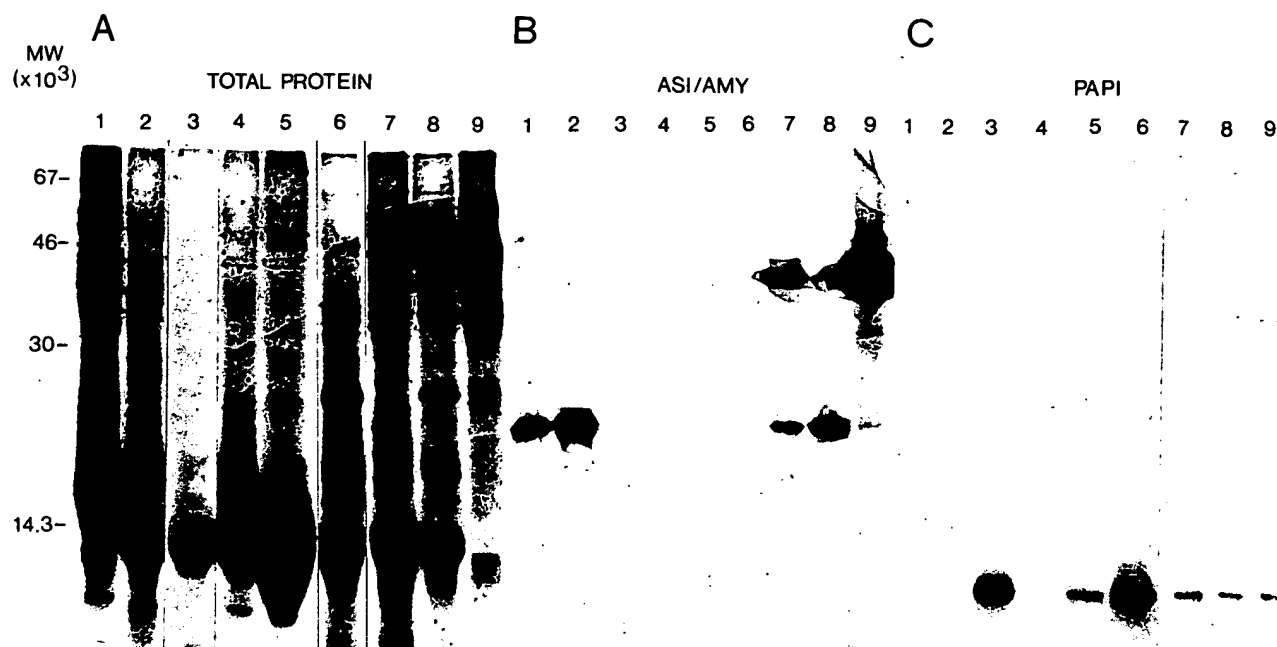


Fig. 5A-C. Immunoprecipitation of PAPI, ASI and  $\alpha$ -amylase from in-vitro translation of mRNA from different tissues of barley grains. Presented are fluorograms from SDS-polyacrylamide gels carrying [ $^{35}$ S]methionine-labeled translation products (Anderson and Blobel 1983). A Total protein products; B immunoadsorptions with antibodies towards ASI (23 kDa) and  $\alpha$ -amylase (44 kDa); C immunoadsorptions with antibodies to PAPI (12 kDa). Tissue samples from which mRNA preparations were obtained:

Lane	Tissue
1	20 dpa starchy endosperm
2	30 dpa starchy endosperm
3	Mature dry endosperm
4	20 dpa aleurone
5	30 dpa aleurone

Lane	Tissue
6	Mature dry aleurone
7	Untreated incubated aleurone
8	ABA-treated incubated aleurone
9	GA <sub>3</sub> -treated incubated aleurone

Panel A was from 2-d exposure; panel B was from a 7-d exposure; panel C, lanes 1-6 were from a 28-d exposure, and lanes 7-9 were from a 32-d exposure

30 dpa (lane 5), or mature dry grain (lane 6), while PAPI mRNA was present beginning at 30 dpa (Fig. 5C, lane 5), and was substantially increased in dry seed (lane 6). PAPI was also present in the endosperm of mature grain (Fig. 5C, lane 3).

In aleurone layers derived from rehydrated grains, the pattern of ASI (and  $\alpha$ -amylase) mRNA abundance reflects the pattern of in-vivo protein synthesis previously described (Mundy 1984). (Since ASI and  $\alpha$ -amylase are tightly associated, antisera to either native protein cross-react with the other.) Both ASI and  $\alpha$ -amylase mRNA are present in untreated aleurone layers (Fig. 5B, lane 7). Treatment with ABA increases ASI and decreases  $\alpha$ -amylase (lane 8), while treatment with GA<sub>3</sub> has an opposite effect (lane 9). The results observed with respect to translatable  $\alpha$ -amylase mRNA are similar to those previously reported (Mozer 1980; Higgins et al. 1982), and demon-

strate that the tissues utilized here responded appropriately to the hormones. In contrast, PAPI is expressed at similar levels regardless of the presence or absence of hormones (Fig. 5C, lanes 7-9).

The quantity of PAPI mRNA in aleurone tissue cultures, relative to levels in mature seed and relative to ASI- and  $\alpha$ -amylase-mRNA levels seems to be substantially lower in this experiment than in either the hybridization experiments (see above) or the in-vivo experiments (below). When similar experiments were performed utilizing wheat-germ extract instead of reticulocyte lysate to translate mRNAs, the abundance of PAPI mRNA relative to ASI and  $\alpha$ -amylase was apparently even less than determined here (data not presented). Thus, the discrepancy in estimates of PAPI mRNA abundance as determined by hybridization and in-vivo labeling versus in-vitro translation of total mRNAs is not dependent upon the translation system. To

Fi;  
Se  
(4.  
gel  
—  
La  
—  
A

B,

The  
lab.  
diff  
resi

—  
 $\alpha$ -A  
ASI  
PAI

exp  
PA  
in-  
in

Syn  
bat

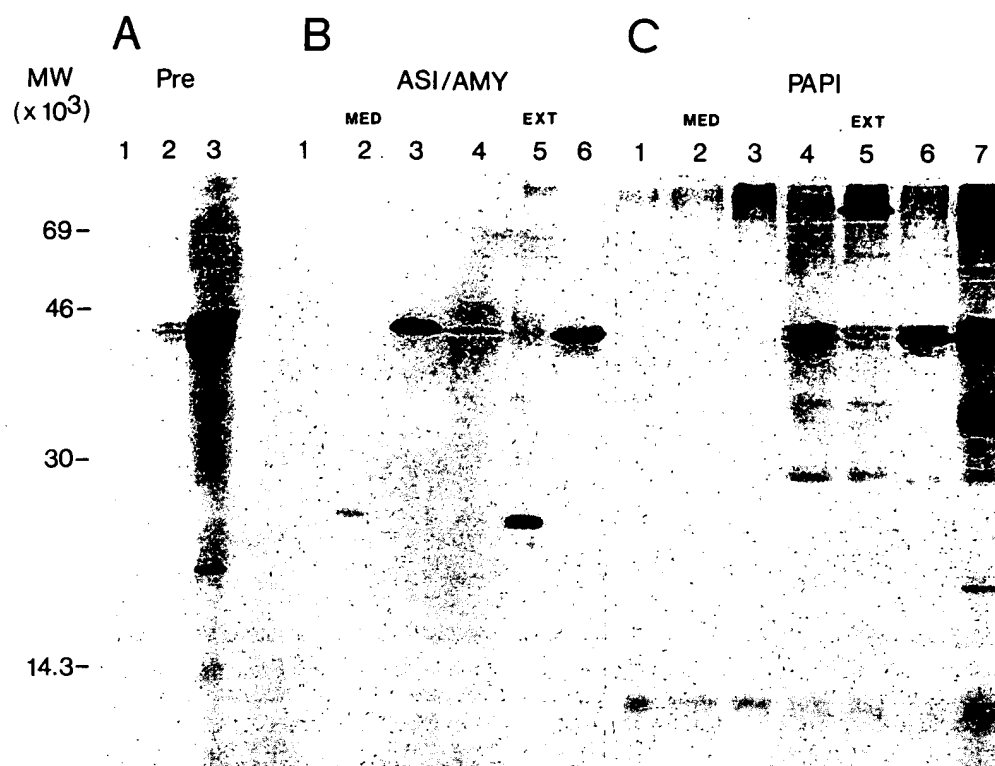


Fig. 6A-C. Immunoprecipitations of PAPI, ASI and  $\alpha$ -amylase from excised aleurone layers incubated in vivo with [ $^{35}$ S]methionine. Sets of 115 aleurone layers from de-embryonated half-seeds were incubated in 22 ml buffer containing  $3.7 \cdot 10^7$  Bq [ $^{35}$ S]methionine ( $4.3 \cdot 10^{13}$  Bq  $\cdot$  mmol $^{-1}$ ) for 20 h in the absence, or presence of GA $_3$  or ABA. Presented are fluorograms of SDS-polyacrylamide gels containing labeled proteins selected by A, preimmune IgG; B ASI and  $\alpha$ -amylase anti-IgG; C PAPI anti-IgG:

Lane	Sample	cpm	Lane	Sample	cpm
A 1	Incubation medium from untreated aleurone	$1 \cdot 10^6$	B, C3	Medium from GA $_3$ -treated aleurone	$1 \cdot 10^6$
2	Tissue extract from untreated aleurone	$1 \cdot 10^6$	4	Tissue extract untreated aleurone	$1 \cdot 10^6$
3	Tissue extract from untreated aleurone	$10 \cdot 10^6$	5	Extract ABA-treated aleurone	$1 \cdot 10^6$
B, C1	Incubation medium from untreated aleurone	$1 \cdot 10^6$	6	Extract GA $_3$ -treated aleurone	$1 \cdot 10^6$
2	Medium from ABA-treated aleurone	$1 \cdot 10^6$	C 7	Tissue extract untreated aleurone	$5 \cdot 10^6$

The fluorogram from panel B was exposed 7 d, while that from C was exposed for 28 d. The relative abundance of the various labeled proteins was estimated by scanning densitometry of the bands on the fluorograms, and the relative amounts in the different immunoprecipitates was calculated by dividing the scan peak area by the (days exposure)  $\times$  (number of methionine residues in the mature protein)  $\times$  ([ $^{35}$ S]methionine decay factor):

	Media			Extract		
	Un	ABA	GA	Un	ABA	GA
$\alpha$ -Amylase	0.04	0.0009	0.15	0.01	0.07	0.4
ASI	0.0001	0.02	0.0003	0.004	0.2	0.002
PAPI	0.08	0.07	0.08	0.05	0.05	0.05

explain this discrepancy, we speculate that the PAPI mRNA may be translated with less efficiency in-vitro in the presence of other mRNAs present in abundance in incubated aleurone tissue.

*Synthesis of PAPI protein in aleurone layers incubated in vivo.* The presence of relatively large

amounts of mRNA for PAPI in mature aleurone layers indicated that PAPI protein would be one of the most abundant proteins synthesized in those tissues after rehydration of the desiccated grain. However, previous studies by others utilizing SDS-PAGE analysis of labeled by incubation of aleurone layers in the presence of radioactive amino

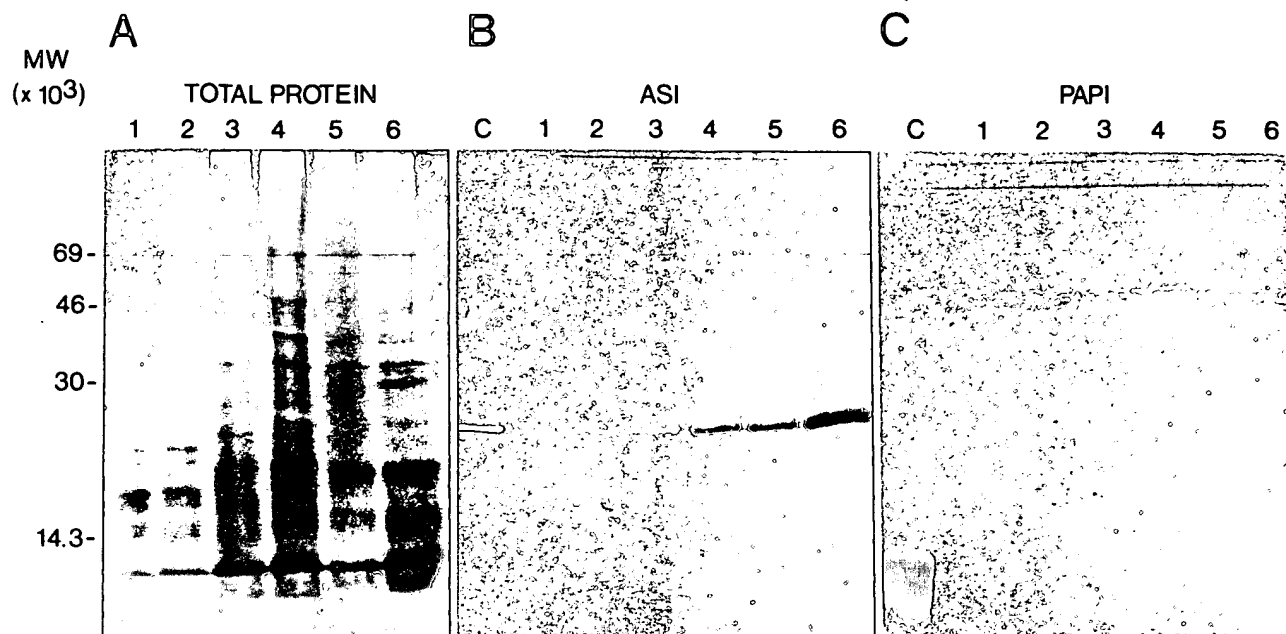


Fig. 7A-C. Immunoblotting assay of ASI and PAPI at different times during development of barley grains. Proteins separated by SDS-PAGE were transferred to nitrocellulose and detected by immunoperoxidase. Panel A presents a silver stain of the total proteins present in each sample (400  $\mu$ g/lane). Panel B presents detection with anti-ASI IgG, and panel C, with anti-PAPI. Lane c represents 1  $\mu$ g of purified protein standard for each. The proteins were extracted from grains 14 (lane 1), 18 (2), 26 (3), 30 (4), 35 (5), and 44 (lane 6) dpa

acids (Mozer 1980; Higgins et al. 1982) had not provided evidence for this. We therefore incubated excised, mature aleurone layers in the presence of [ $^{35}$ S]methionine with, or without, GA<sub>3</sub> or ABA, and the presence of newly synthesized, labeled PAPI was assessed by immunoprecipitation and SDS-PAGE autoradiography. These experiments were complicated by two factors: (1) mature PAPI proteins contain only a single methionine residue, and (2) relatively large amounts of PAPI protein are contained in mature aleurone layers (see below). Thus, newly synthesized protein is diluted out by already existing protein during immunoprecipitation. In order to circumvent the latter problem, we tried varying concentrations of tissue extract (antigen) in immunoabsorption experiments. Extract concentrations of  $0.25 \cdot 10^6$ ,  $1 \cdot 10^6$ ,  $5 \cdot 10^6$ ,  $10 \cdot 10^6$  and  $20 \cdot 10^6$  cpm per 600  $\mu$ l reaction volume (specific activity  $1 \cdot 10^4$ – $1.2 \cdot 10^4$  cpm  $\cdot \mu$ g<sup>-1</sup> protein) were used to optimize the ratio of antigen to antibody. These results demonstrated that the intensity of the immunoprecipitated PAPI band, relative to other, nonspecific bands, was inversely correlated with the quantity of extract added to the standard amount of antiserum (data not shown). Thus it is likely that even the smallest amount of extract still represented a condition of antigen excess. For clarity in the figures, only selected sets of results

are presented. In addition, to optimize the amount of protein available for antigen binding in the extracts, we included SDS in our extraction and media resolubilization buffers. Although this denaturing treatment aids solubilization and antigen recognition of certain proteins in freeze-dried samples, it sometimes produces the fairly high levels of nonspecific adsorption of sample polypeptides to antibodies and immunoabsorbent seen here (Fig. 6A, lanes 1–3). However, comparison of these lanes to lanes containing proteins adsorbed by specific antiserum clearly shows that a protein of 10 kDa (PAPI) is specifically recognized.

Again, consistent with previous in-vivo experiments (Mundy et al. 1984), and with the results obtained with mRNA translation in vitro (above), ABA causes a decrease in  $\alpha$ -amylase and an increase in ASI in tissue extract (Fig. 6B, lane 5), while GA<sub>3</sub> has the opposite effect (lane 6). Gibberellic acid also causes the appearance of labeled  $\alpha$ -amylase in the culture medium (lane 3). In contrast, PAPI is found in the culture medium in levels that are little influenced by the hormones (Fig. 6C, lanes 1–3). Similar results are seen in tissue-extract samples (Fig. 6C, lanes 4–6). When the amount of tissue-extract sample used for immunoabsorption was increased fivefold (Fig. 6C, lane 7), the amount of specifically adsorbed PAPI (relative to



the background pattern of nonspecifically adsorbed proteins) decreased substantially. Increases of 20-fold in quantity of tissue extract over that used in lanes 4–6 yielded patterns that were not different from those obtained with preimmune serum (data not presented).

We utilized these data to estimate the relative quantities of the three proteins synthesized under the different conditions (Fig. 6, legend). These calculations ignore the effect of unlabeled protein on efficiency of immunoprecipitation, and thus will substantially underestimate the amount of PAPI. In addition, the relative values cannot be compared between one set and another (for example, between the control and GA<sub>3</sub>-treated sets) because the amino-acid pool sizes may differ substantially. Nevertheless, these calculations demonstrate that PAPI is the most abundant of the three proteins in the media samples from control and ABA-treated aleurone, and amylase is apparently only twice as abundant in incubation medium from GA<sub>3</sub>-treated tissue. A similar pattern is obtained with the tissue extracts, except that ASI is the most abundant protein of the three in ABA-treated tissue. It is important to note that the relative ratios of newly synthesized PAPI to  $\alpha$ -amylase in this experiment in control and GA<sub>3</sub>-treated aleurone layers are very similar to the relative ratios of hybridizable RNA measured in Fig. 4. This concordance of results from two independent experiments supports our assertion that the apparently low levels of translatable PAPI mRNA observed in Fig. 5 are likely to be an artefact.

*Steady-state levels of PAPI and ASI protein assessed by immune blotting.* Immune blotting was used to assess the relative amounts of ASI and PAPI proteins in developing whole barley grains (Fig. 7). It can be seen that the 20-kDa ASI protein is already detectable at 14 dpa (Fig. 7B, lane 1), and that its steady-state abundance progressively increases to the level detected in the 44-dpa mature sample (Fig. 7B, lane 6). In contrast, the 10-kDa mature PAPI protein is first detected in this experiment at 35 dpa (Fig. 7C, lane 5), and increases to the level present at 44 dpa (lane 6). The larger band detected in the PAPI control sample (Fig. 7C, lane C) may represent a small amount of preprotein present in the purified sample; a similar minor band of approx. 14 kDa was observed in some tissue-protein analyses (data not presented). In other, similar experiments (data not presented), the relative abundance of PAPI in aleurone layers and endosperm from rehydrated mature grains was similar to that detected in extracts from mature grain.

## Discussion

We have described a cloned cDNA representing a barley protein (PAPI) with homologies to a known amylase inhibitor, and to trypsin-inhibitory domains of Bowman-Birk type bifunctional protease inhibitors. However, to date, the PAPI protein has been shown not to inhibit the action of trypsin, chymotrypsin, subtilisin, crude *Aspergillus niger* protease, and papain on synthetic and protein substrates, nor has it inhibited  $\alpha$ -amylases from human saliva, hog pancreas, *Tenebrio molitor*, barley malt, *Aspergillus niger* or *Bacillus subtilis* on soluble starch (Svensson et al. 1986). The determination of the target enzyme(s), if any, for PAPI, will take both time and perhaps luck, considering the high specificity of protease and/or amylase inhibitor interactions (Laskowski and Kato 1980; Campos and Richardson 1983; Mundy et al. 1983; Odani et al. 1983).

The importance of the protein and cloned cDNA presented here lies more in the tissue specificity and abundance of its expression. We have demonstrated that, in contrast to ASI which is expressed primarily, if not exclusively, in endosperm tissue temporally corresponding to the peak of storage-protein synthesis (Mathews and Mifflin 1980; Giese and Hejgaard 1984; Mundy et al. 1986), PAPI is expressed late during grain development. Its abundant expression in the aleurone cells of mature seeds, a layer which comprises only a small fraction of the grain, and its abundant expression in aleurone layers isolated after rehydration of desiccated grains lead us to speculate that it may be targeted there to inhibit exogenous amylase and/or protease activity.

In contrast to the effects of ABA and GA<sub>3</sub> on the synthesis of  $\alpha$ -amylase and ASI in excised barley aleurone layers, the synthesis of PAPI, measured in vitro (Fig. 5), and in vivo (Fig. 6), is not influenced by incubation with either hormone. Because the effects of these hormones on overall protein synthesis in aleurone layers is profound (Ho and Varner 1976; Mozer 1980), the apparent constitutive synthesis of PAPI in these cells indicates that the protein plays a role in maintenance of tissue integrity. A similar pattern of synthesis has been shown for the barley chymotrypsin-inhibitor 1 protein (Mundy et al. 1986), a result which lends support to our speculation that germinating aleurone cells constitutively synthesize proteinase inhibitors as some sort of protective response. Interestingly, at least some of the de-novo synthesized PAPI can be identified in the media of cultured layers (Fig. 6), indicating that the protein is se-

creted by aleurone cells. That PAPI is secreted is consistent with the presence of a signal peptide as determined from the cDNA sequence.

Experiments quantitating the relative abundance of hybridizable PAPI RNA (Fig. 4), and the quantity of PAPI protein synthesized *in vivo* (Fig. 6), both indicate that PAPI mRNA and protein are expressed at levels comparable to  $\alpha$ -amylase in  $GA_3$ -treated aleurone tissue. Experiments assessing PAPI mRNA levels, as judged by ability to produce PAPI protein in a cell-free protein-synthesis system (Fig. 5), are however at variance with these results. We choose to favor the agreement displayed by two different measurements, and suggest that the latter results may be the consequence of some selective disadvantage PAPI mRNA has in cell-free translation systems relative to other aleurone mRNAs. This remains to be proved. However, this observation, coupled with the paucity of methionine residues in the protein, might explain why other studies utilizing cell-free translation of RNA from excised aleurone layers (Mozer 1980; Higgins et al. 1982) also did not identify a prominent 12-kDa protein product.

Both hybridization (Fig. 3B) and translation-immunoprecipitation (Fig. 5C) experiments indicate that PAPI mRNA must be one of the most abundant mRNAs late during aleurone cell development and in mature, dry seeds. As demonstrated by the disappearance of ASI mRNA in mature endosperm (Fig. 5B), most mRNAs specifying storage and other proteins synthesized during the peak of endosperm development are lost as the seed matures (Mathews and Mifflin 1980; Giese and Hejgaard 1984; Mundy et al. 1986). This abundance of PAPI mRNA in the dry seed raises the question of whether the PAPI mRNA present in aleurone layers after rehydration is actively synthesized or whether it represents stored mRNA synthesized prior to desiccation. This question can be best addressed by runoff transcription experiments utilizing barley-aleurone protoplast nuclei (Jacobsen and Beach 1985). The PAPI sequence shares one feature with genes that are expressed in aleurone cells after rehydration but not in developing seed, namely the two  $\alpha$ -amylase (Rogers 1985) and the aleurain thiol protease (Rogers et al. 1985) genes, and that is a high (64%) G + C content. In this regard it differs substantially from two genes that are expressed in endosperm during the peak of storage-protein synthesis; hordein (Rasmussen et al. 1983) storage-protein and protein-Z (Hejgaard et al. 1985) genes both have about 50% G + C in their coding sequences. Whether the high G + C content of PAPI and the  $GA_3$ -induced aleu-

rone genes, and the complex secondary structures that result (Rogers 1985; Rogers et al. 1985) have any role in mRNA stability is as yet unclear.

Our results also provide additional data regarding the expression of ASI. This protein is presumably synthesized in endosperm cells to protect the seed proteins against microbial attack and to limit the degradation of starch granules by the plant's own  $\alpha$ -amylase. This enzyme activity must be carefully regulated in cereals of temperate zones which are often stimulated to germinate precociously by wet weather. By inhibiting low levels of amylase activity secreted into the endosperm, ASI may act as one of many factors which mediate the change from dormancy to germination in these seeds. Interestingly, ASI is not synthesized in developing aleurone tissue (Fig. 5), but its synthesis can be enhanced in mature, excised aleurone layers by ABA and reduced by  $GA_3$  (Fig. 5). Such a pattern of expression and hormonal control in mature aleurone layers is not seen for several other seed proteins tested to date (Mundy et al. 1986). This indicates that the ABA induction of ASI synthesis and of a limited number of other polypeptides is specific and may be related to their physiological functions. Although the general mechanism of ABA control of protein synthesis in cultured layers is not known (Ho and Varner 1976), its effect on ASI probably occurs at the level of mRNA abundance. Absciscic acid has been shown to prevent, at least in part, the increase in levels of mRNAs encoding  $\alpha$ -amylase in  $GA_3$ -treated mature aleurone layers (Mozer 1980; Higgins et al. 1982). In addition, type A or, more properly,  $\alpha$ -amylase 1 (Svensson et al. 1985) mRNA is expressed at substantial levels in untreated mature aleurone layers (Rogers 1985), and ABA treatment causes a substantial decrease in this base-line expression (Fig. 3). Taken together, these results support the idea that ASI synthesized in aleurones represents an additional co- or post-translational control of endogenous  $\alpha$ -amylase activity that is promoted by the same hormone that itself causes a decrease in  $\alpha$ -amylase mRNA.

This work was supported in part by grant 83-CRCR-1-1332 from the U.S. Department of Agriculture.

## References

- Anderson, D.J. Blobel, G. (1983) Immunoprecipitation of proteins from cell-free translations. *Methods Enzymol.* **96**, 111-120
- Boisen, S., Andersen, C.Y. Hejgaard, J. (1981) Inhibitors of chymotrypsin and microbial serine proteases in barley grains. *Physiol. Plant.* **52**, 167-176

- Bollum, F.J. (1975) Antibody to terminal deoxynucleotidyl transferase. *Proc. Natl. Acad. Sci. USA* **72**, 4119-4122
- Campos, F.A.P., Richardson, M. (1983) The complete amino acid sequence of the bifunctional  $\alpha$ -amylase inhibitor from seeds of ragi (Indian finger millet, *Eleusine coracana* Gaertn.) *FEBS Lett.* **152**, 300-304
- Campos, F.A.P., Richardson, M. (1984) The complete amino acid sequence of the  $\alpha$ -amylase inhibitor I-2 from seeds of ragi (Indian finger millet, *Eleusine coracana* Gaertn.). *FEBS Lett.* **167**, 221-225
- Croy, R.R., Lycett, G., Gatehouse, J.A., Yarwood, J.N., Boulter, D. (1982) Cloning and analysis of cDNAs encoding plant storage protein precursors. *Nature* **295**, 76-78
- Dayhoff, M.O., Barker, W.C., Hunt, L.T. (1983) Establishing homologies in protein sequences. *Methods Enzymol.* **91**, 524-545
- Giese, H., Hejgaard, J. (1984) Synthesis of salt-soluble proteins in barley. Pulse-labeling study of grain filling in liquid-cultured detached spikes. *Planta* **161**, 172-177
- Giese, H., Hopp, H.E. (1984) Influence of nitrogen nutrition on the amount of hordein, protein Z, and  $\beta$ -amylase messenger RNA in developing endosperms of barley. *Carlsberg Res. Commun.* **49**, 365-383
- Hejgaard, J., Rasmussen, S.K., Brandt, A., Svendsen, I. (1985) Sequence homology between barley endosperm protein Z and protease inhibitors of the  $\alpha$ 1-antitrypsin family. *FEBS Lett.* **180**, 89-94
- Hejgaard, J., Svendsen, I., Mundy, J. (1983) Barley  $\alpha$ -amylase/subtilisin inhibitor. II. N-terminal amino acid sequence and homology with inhibitors of the soybean trypsin inhibitor (Kunitz) family. *Carlsberg Res. Commun.* **48**, 91-94
- Higgins, T.J.V., Jacobsen, J.V., Zwar, J.A. (1982) Gibberellic acid and abscisic acid modulate protein synthesis and mRNA levels in barley aleurone layers. *Plant Mol. Biol.* **1**, 191-215
- Ho, D.T.H., Varner, J.E. (1976) Response of barley aleurone layers to abscisic acid. *Plant Physiol.* **57**, 175-178
- Jacobsen, J.V. (1984) Regulation of protein synthesis in aleurone cells by gibberellins and abscisic acid. In: *The biochemistry and physiology of gibberellins*, vol. 2, pp. 159-187, Crozier, A., ed. Praeger, New York
- Jacobsen, J.V., Beach, L.R. (1985) Control of transcription of  $\alpha$ -amylase and rRNA genes in barley aleurone protoplasts by gibberellin and abscisic acid. *Nature* **316**, 275-277
- Jensen, S.A., Munck, L., Martens, H. (1982) The botanical constituents of wheat and wheat milling fractions. I. Quantification by autofluorescence. *Cereal Chem.* **59**, 477-484
- Jonassen, I. (1980) Characteristics of hiproly barley II. Quantification of two proteins contributing to its high lysine content. *Carlsberg Res. Commun.* **45**, 59-68
- Jonassen, I., Ingversen, J., Brandt, A. (1981) Synthesis of SP II albumin,  $\beta$ -amylase and chymotrypsin inhibitor CI-1 on polysomes from the endoplasmic reticulum of barley endosperm. *Carlsberg Res. Commun.* **46**, 175-181
- Joubert, F.J., Kruger, H., Towshend, G.S., Botes, D.P. (1979) Purification, some properties and the complete primary structures of two protease inhibitors (DE-3 and DE-4) from *Macrotyloma axillare* seed. *Eur. J. Biochem.* **97**, 85-91
- Laskowski, M., Kato, I. (1980) Protein inhibitors of proteinases. *Annu. Rev. Biochem.* **49**, 593-696
- Mathews, J.A., Mifflin, B.J. (1980) In-vitro synthesis of barley storage proteins. *Planta* **149**, 262-268
- Maxam, M.A., Gilbert, W. (1980) Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* **65**, 497-560
- Mozer, T.J. (1980) Control of protein synthesis in barley aleurone layers by the plant hormones gibberellic acid and abscisic acid. *Cell* **20**, 479-485
- Mundy, J. (1984) Hormonal regulation of  $\alpha$ -amylase inhibitor synthesis in germinating barley. *Carlsberg Res. Commun.* **49**, 439-444
- Mundy, J., Brandt, A., Fincher, G. (1985) Messenger RNAs from the scutellum and aleurone of germinating barley encode (1-3, 1-4)- $\beta$ -D-glucanase,  $\alpha$ -amylase, and carboxypeptidase. *Plant Physiol.* **79**, 867-871
- Mundy, J., Hejgaard, J., Hansen, A., Hallgren, L., Jorgensen, K.G., Munck, L. (1986) Differential synthesis in vitro of barley aleurone and starchy endosperm proteins. *Plant Physiol.* (in press)
- Mundy, J., Hejgaard, J., Svendsen, I. (1984) Characterization of a bifunctional wheat inhibitor of endogenous  $\alpha$ -amylase and subtilisin. *FEBS Lett.* **167**, 210-214
- Mundy, J., Svendsen, I., Hejgaard, J. (1983) Barley  $\alpha$ -amylase/subtilisin inhibitor. I. Isolation and characterization. *Carlsberg Res. Commun.* **48**, 81-90
- Odani, S., Koide, T., Ono, T. (1983) The complete amino acid sequence of barley trypsin inhibitor. *J. Biol. Chem.* **258**, 7998-8003
- Ponz, F., Paz-Ares, J., Hernandez-Lucas, C., Carbonero, P., Garcia-Olmedo, F. (1983) Synthesis and processing of thionin precursors in developing endosperm from barley (*Hordeum vulgare* L.). *EMBO J.* **2**, 1035-1040
- Rasmussen, S.K., Hopp, H.E., Brandt, A. (1983) Nucleotide sequences of cDNA clones for B1 hordein polypeptides. *Carlsberg Res. Commun.* **48**, 187-199
- Richardson, M. (1979) The complete amino acid sequence and the trypsin reactive (inhibitory) site of the major proteinase inhibitor from the fruits of aubergine (*Solanum melongena* L.). *FEBS Lett.* **104**, 322-326
- Rogers, J.C. (1985) Two barley  $\alpha$ -amylase gene families are regulated differently in aleurone cells. *J. Biol. Chem.* **260**, 3731-3738
- Rogers, J.C., Dean, D., Heck, G.R. (1985) Aleurain: a barley thiol protease closely related to mammalian cathepsin H. *Proc. Natl. Acad. Sci. USA* **82**, 6512-6516
- Rogers, J.C., Milliman, C. (1983) Isolation and sequence analysis of a barley  $\alpha$ -amylase cDNA clone. *J. Biol. Chem.* **258**, 8169-8174
- Shewry, P.R., Lafandra, D., Salcedo, G., Aragoncillo, C., Garcia-Olmedo, F., Lew, E. J.-L., Dietler, M.D., Kasarda, D.D. (1984) N-terminal amino acid sequences of chloroform/methanol-soluble proteins and albumins from endosperms of wheat, barley and related species. *FEBS Lett.* **175**, 359-363
- Shivaraj, B., Pattabiraman, T.N. (1981) Natural plant enzyme inhibitors: Characterization of an unusual  $\alpha$ -amylase/trypsin inhibitor from ragi (*Eleusine coracana* Gaertn.). *Biochem. J.* **193**, 29-36
- Svensson, B., Asano, K., Jonassen, I., Poulsen, F., Mundy, J., Svendsen, I. (1986) Isolation and characterization of a barley seed protein with M<sub>r</sub> 10,000 showing homology with an  $\alpha$ -amylase inhibitor for Indian Finger Millet. *Carlsberg Res. Commun.* (in press)
- Svensson, B., Mundy, J., Gibson, R.M., Svendsen, I. (1985) Partial amino acid sequences of  $\alpha$ -amylase isozymes from barley malt. *Carlsberg Res. Commun.* **50**, 15-22
- Towbin, H., Staehelin, T., Gordon, J. (1979) Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc. Natl. Acad. Sci. USA* **76**, 4350-4354
- Weslake, R.J., MacGregor, A.W., Hill, R.D. (1983) An endogenous  $\alpha$ -amylase inhibitor in barley kernels. *Plant Physiol.* **72**, 809-812

Received 20 January; accepted 18 March 1986

**Claim 63**  
**Lipoxygenase****Isolation and characterization  
of a soybean (*Glycine max*) lipoxygenase-3 gene**

Richard L. Yenofsky, Miriam Fine, and Cecilia Liu\*

Phytogen, 101 Waverly Drive, Pasadena, CA 91105, USA

**Summary.** cDNA and genomic DNA clones for a soybean (*Glycine max*) seed lipoxygenase-3 (lox-3) isozyme have been isolated and characterized by molecular cloning and sequence analysis. The cDNA clone p3H9 was identified as encoding a lipoxygenase isozyme based on hybridization selection and immunoprecipitation assays. The insert from p3H9 was used to screen 700,000 recombinants in a phage genomic library. Two overlapping genomic clones were isolated. After localizing the mRNA coding region to the mid-portion of one of these clones, nucleotide sequence analysis was performed. The structure of the gene was deduced following the additional cloning and sequencing of nearly full length cDNAs. Primer extension and S1 protection experiments were used to define the 5' boundary of the gene. The gene is encoded by nine exons interrupted by eight introns. All exon-intron junctions conform to the GT-AG rule. The deduced amino acid sequence consists of 857 amino acids with a molecular weight of 96,663 daltons. The 5' untranslated segment contains approximately 46 nucleotides before the first methionine (ATG) codon. A typical TATA box occurs some 30 nucleotides upstream of the message start site while a CAAG sequence occurs in place of a CAAT box some 80 nucleotides upstream of the start of transcription. Four polyadenylation signals are present in the 3' untranslated region. At least two of these are used, giving rise to cDNA clones bearing 3' untranslated segments of 148 and 196 nucleotides. To identify which member of the lox gene family had been isolated, a gene-specific probe from the 3' untranslated region of p3H9 was hybridized to dot blots of soybean embryo RNAs derived from lox-1, -2, or -3 null mutants. No hybridization to RNA from the lox-3 null mutant was seen, while the remaining mutant RNAs showed no reduction in hybridization. Further confirmation that the gene encodes a lox-3 isozyme was obtained by a comparison of deduced and determined amino acid sequence data. An analysis of the lox-3 sequence and a comparison to other available soybean lox sequences is also presented.

**Key words:** cDNA and genomic clones – *Glycine max* – Lipoxygenase – Soybean seed protein

\* Present address: Department of Neural Science, City of Hope Hospital, 4050 Duarte Road, Duarte, CA 91010, USA

Offprint requests to: R. Yenofsky

**Introduction**

Lipoxygenases (linoleate: oxygen oxidoreductase, EC 1.13.11.12) are a widely distributed group of nonheme iron-containing enzymes that catalyze the oxygenation of polyunsaturated fatty acids or their esters containing a *cis*, *cis*-1,4-pentadiene system to form monohydroperoxides as primary products. At least three lipoxygenases have been identified in mammalian systems and they are the focus of intensive interest due to their role in the biosynthesis of prostaglandins, leukotrienes and lipoxins (for reviews, see Samuelsson 1983; Needleman et al. 1986). On the other hand, the physiological role of plant lipoxygenases (lox) is poorly understood. The enzyme is thought to play a role in fruit ripening and abscission (Veldink et al. 1977), senescence (Leshem 1984), and resistance to plant pathogens (Lupu et al. 1980; Hildebrand et al. 1986; Ocampo et al. 1986). Lox has been the focus of strong commercial interest because its activity in a variety of foods can result in the generation of objectionable as well as pleasant flavors arising from aromatic aldehydes and alcohols generated secondarily from lox-produced hydroperoxides (for reviews, see Axelrod 1974; Eskin et al. 1977; Galliard and Chan 1980). From a commercial viewpoint the negative and positive aspects of lox activity in a major crop such as soybean are of particular interest. On the negative side, the rancidity of soy oil and the poor quality of soy protein are thought to arise as a consequence of lox activity (Mustakas et al. 1969; Wolf 1975; Hildebrand and Kito 1984). On the positive side, defatted soy extract is an important component in bread-making. Lox, which is the active ingredient in the extract, is responsible for two processes: (i) the bleaching of carotenoid pigments, resulting in a whiter bread, and (ii) reactions involving oxidation and subsequent crosslinking of wheat glutens, resulting in enhanced texture (for review, see Eskin et al. 1977).

Four lox isozymes have been identified in soybean seeds (Axelrod et al. 1981), and as a group they constitute 1%–2% of the total seed protein (Vernooy-Gerritsen et al. 1983). The enzyme consists of a single polypeptide chain of roughly 95,000 daltons. The isozymes have been divided into 2 classes or types that differ in substrate preference, heat stability, pH optimum, and response to calcium. The enzyme crystallized by Theorell et al. (1947) is a type-1 enzyme and has been designated lox-1. It is heat stable, has

a pH optimum of about 9.0, and prefers anionic substrates (fatty acids). The type-2 enzymes, lox-2 (Christopher et al. 1970) and lox-3 (Christopher et al. 1972), are more easily inactivated by heat, have pH optima closer to neutrality and prefer esterified substrates. Lox-2 activity is stimulated by calcium, while lox-3 is inhibited. Lox-3 can be chromatographically resolved into two fractions, designated 3A and 3B, whose properties are virtually indistinguishable. Additional differences also exist between the isozymes. Lox-3 generates singlet oxygen with higher efficiency than lox-1 or lox-2 (Kanofsky and Axelrod 1986). The enzymes also differ in the proportions of 9- and 13-hydroperoxides formed. Lox-1 shows a preference for the 13 position, while lox-2 and lox-3 utilize the 9 or 13 positions in roughly equal proportions (Christopher and Axelrod 1971).

The involvement of lox in determining the quality of soybean products makes it an interesting candidate for study. As an approach to understanding how lox is regulated, we have undertaken the molecular cloning and analysis of a soybean lipooxygenase gene. Recently, the complete amino acid sequence of soybean lox-1 was deduced from its corresponding cDNA (Shibata et al. 1987). A partial sequence for a soy "lox-1-like" protein has also been reported (Start et al. 1986). Here, we present the first complete nucleotide sequence and structure of a soy lox-3 gene.

## Materials and methods

**Isolation of size-fractionated poly(A)<sup>+</sup> RNA.** Polysomes from mid-maturation stage soybean embryos (*Glycine max*; variety Dare) were isolated as described by Goldberg et al. (1981). After pelleting through a cushion of CsCl, the polyosomal RNA was enriched for poly(A)<sup>+</sup> mRNA by adsorption to oligo (dT) cellulose. The eluted material was subjected to zone centrifugation at 15° C in the Spinco SW41 rotor through a linear 5%–30% sucrose gradient in 20 mM Tris (pH 7.6), 10 mM NaCl and 0.1% SDS. Centrifugation was for 11 h at 36,000 rpm. Fractions were collected by precipitation in ethanol, then resuspended in H<sub>2</sub>O. Aliquots of each fraction were translated in a reticulocyte in vitro translation system as described by the manufacturer (Promega) and the translation products separated by one-dimensional denaturing polyacrylamide gel electrophoresis. Fractions showing substantial enrichment of poly(A)<sup>+</sup> mRNA encoding for proteins greater than 90,000 daltons were pooled and used for cloning.

**Construction and screening of soybean embryo cDNA libraries.** Construction of the initial cDNA library from size-fractionated poly(A)<sup>+</sup> RNA followed procedures described previously (Yenofsky et al. 1982). The library was screened according to the procedure of Grunstein and Hogness (1975) with a cDNA probe made from the same RNA that was used to construct the library. Replicas of the library were hybridized with the cDNA probe for two different time periods (3 and 17 h) in order to be able to distinguish between clones derived from abundant and less abundant messages in the labeled cDNA population. Clones exhibiting rapid hybridization with the labeled probe were selected for further analysis.

To obtain full- or nearly full-length cDNA clones, a second library was made from total polysomal poly(A)<sup>+</sup> RNA according to Gubler and Hoffman (1983), except that the double-stranded cDNA was first size-fractionated to

enrich for cDNAs greater than 1,500 base pairs before joining it with the vector pBR322. This library was screened with the insert isolated from p3H9, a partial-length lox clone obtained from the first library. Approximately 2,000 transformants were screened, and 16 reacted with the insert. Restriction analysis indicated that 13 were homologous to p3H9. Of these, 4 were close to full-length.

**Hybrid selection, immunoprecipitation and RNA blot procedures.** RNA denaturation, fractionation and transfer to nitrocellulose, as well as hybridization selection, were carried out as previously described (Yenofsky et al. 1982) except that the incubation temperature for hybrid selection was 50° C. Immunoprecipitation of in vitro translation products with *Staphylococcus aureus* Cowan 1 strain cells was performed as described by the manufacturer (The Enzyme Center, Boston, Mass.). RNA was dotted onto Zeta-Probe membranes following the manufacturer's protocol (Bio-Rad).

**Isolation of lipooxygenase genomic clones.** A "HaeIII-AluI partial" soybean genomic library, provided by R. Goldberg, was screened by plaque hybridization with the isolated lox cDNA insert from clone p3H9, using standard conditions (Maniatis et al. 1982).

**DNA sequencing.** Labeled genomic DNA fragments were sequenced by the method of Maxam and Gilbert (1980). Over 90% of the genomic sequence was determined from both strands. cDNA clones were sequenced using both chemical degradation and dideoxy chain termination methods (Sanger et al. 1977). When double-stranded plasmid DNA derived from minilysates was used as the source of material for dideoxy sequencing, we encountered severe background problems. To overcome this, minilysates were pretreated with Bal 31 nuclease for 10 min at 30° C, under conditions described by the manufacturer (New England Biolabs). The reaction was terminated by addition of EGTA to a final concentration of 20 mM and the DNA was phenol extracted and ethanol precipitated. The Bal 31 does not affect closed circular DNA, while degrading other nucleic acids in the minilysate.

**Primer extension and S1 protection.** For the primer extension experiment, a 106-nucleotide-long *FspI*-*HaeIII* genomic fragment (nucleotides 920–1025, Fig. 5) was kinase labeled at the *HaeIII* site, mixed with 25 µg of total polysomal RNA or yeast tRNA in 25 µl of 40 mM PIPES (pH 6.4), 400 mM NaCl, 80% formamide, 1 mM EDTA, denatured and allowed to anneal for 3 h at 42° C. The reaction mix was diluted with an equal volume of ice-cold 10 mM Tris (pH 8.0), 1 mM EDTA, collected by ethanol precipitation and resuspended in 50 µl of 50 mM Tris (pH 8.3), 50 mM KCl, 1 mM dithiothreitol (DTT), 10 mM MgCl<sub>2</sub>, 2 mM of each of the four dNTPs, and 100 units of reverse transcriptase. The reaction was incubated at 42° C for 100 min before being terminated by ethanol precipitation, after which the products were resuspended in loading buffer and analyzed by gel electrophoresis.

For S1 protection studies, a 346-nucleotide-long *AhaIII*-*FspI* fragment (nucleotides 579–919, Fig. 5) was kinase labeled at the *FspI* site, mixed with total polysomal RNA or yeast tRNAs, denatured and allowed to anneal as described above. The reactions were terminated by diluting

into 300  $\mu$ l 0.28 M NaCl, 50 mM sodium acetate (pH 4.5), 4 mM ZnSO<sub>4</sub>. Twenty-five hundred units of S1 nuclease were added and the reaction was incubated at 37° C for 30 min, at which time an additional 500 units of S1 were added for another 20 min. The reaction mix was phenol extracted prior to collection by ethanol precipitation, and the products were analyzed as above.

## Results and discussion

### Isolation of cDNA clones

Poly(A)<sup>+</sup> RNA was obtained from mid-maturation stage soybean embryos and then size-fractionated to enrich for lipoxygenase mRNA. A cDNA library was constructed with this enriched RNA and subsequently screened with a cDNA probe made from the same RNA. Clones were chosen based on their ability to hybridize rapidly with the enriched probe. Positively reacting colonies whose plasmids hybridized to mRNAs of appropriate size (greater than 2.8 kb) on Northern blots were chosen for hybrid selection experiments. Three clones were isolated by this screening strategy. The clone with the largest insert (approximately 580 bp), designated p3H9, hybridized to an mRNA slightly less than 3,000 nucleotides in length (Fig. 1, lane 1) and was able to hybrid select an mRNA encoding a protein which comigrated with purified lox (Fig. 2, lane 3).

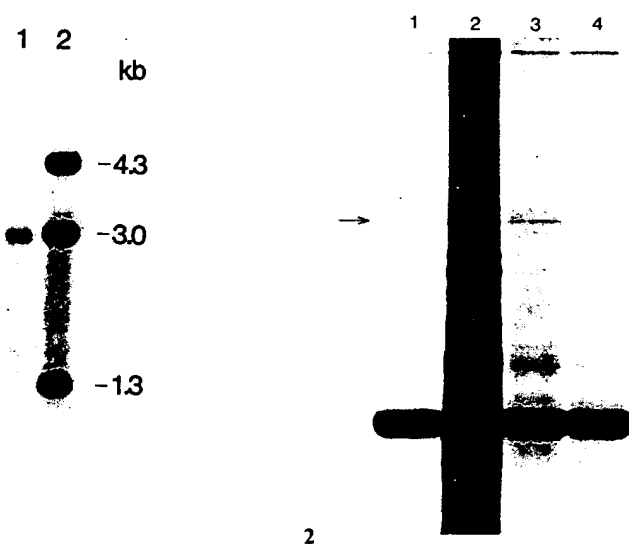


Fig. 1. Electrophoretic behavior of mRNA species homologous to recombinant plasmid p3H9. RNA from soybean embryos was electrophoresed under denaturing conditions through an agarose gel and subjected to Northern blot analysis in the presence of nick-translated p3H9. Lane 1, 5  $\mu$ g soybean total polysomal RNA; lane 2, 3 ng pBR322 molecular weight markers. The numbers refer to the size of the markers, in kilobases.

Fig. 2. Translation products derived from hybrid selected mRNA. Soybean embryo RNA was translated in a reticulocyte lysate in vitro translation system in the presence of <sup>35</sup>S-methionine and the products were electrophoretically separated in denaturing polyacrylamide gels followed by fluorography. Lane 1, no added RNA; lane 2, total polysomal RNA; lane 3, RNA hybrid selected by cDNA clone p3H9; lane 4, RNA hybrid selected by pBR322 control. The arrow indicates the migration position of a purified soybean lipoxygenase marker.

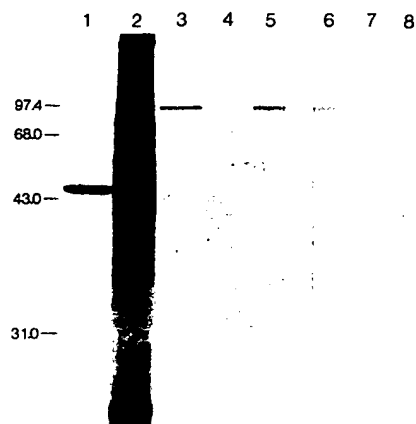


Fig. 3. Characterization of affinity-purified lipoxygenase antibody and immunoprecipitation of hybrid selection products. Reticulocyte lysates were programmed with either total polysomal RNA (lanes 1 through 5) or RNA obtained through hybrid selection (lanes 6–8). Lanes 1 and 2, as in Fig. 2. Translation products in lanes 3–8 were immunoprecipitated by affinity purified lox antibody alone (lanes 3 and 6) or in the presence of excess unlabeled lox (lanes 4 and 7), or urease (lanes 5 and 8). The position of molecular size markers is indicated on the side, in kilodaltons.

To further characterize the clone, we obtained an affinity-purified lox antibody from T. Kaye Peterman. This antibody was able to specifically bind the lox synthesized in a reticulocyte in vitro translation system programmed with total soybean embryo mRNA (Fig. 3, lane 3). Further, the binding of labeled lox by the antibody could be competed out by incubating the lysate in the presence of excess unlabeled lox (Fig. 3, lane 4), but not by urease (Fig. 3, lane 5), a protein found in soybean embryos, the size of which is similar to lox. When hybrid selected RNA was translated in vitro and subjected to the same immunoprecipitation experiment, identical results were obtained (Fig. 3, lanes 6–8). On the basis of these results, we concluded that clone p3H9 encodes a lox isozyme.

In order to determine which lox cDNA had been isolated, total RNAs from mid-maturation stage soybean embryos of mutants lacking either lox-1, -2, or -3 were probed with p3H9/3', a subclone of p3H9 that contains only the gene-specific 3' untranslated region of the cDNA (Fig. 4). No hybridization was observed with the RNA from the mutant lacking lox-3, while hybridization to the RNAs from the mutants lacking lox-1 or lox-2 was equal to that of the RNA from the wild type, suggesting that the gene encodes a lox-3 isozyme.

### Isolation of genomic clones

Genomic clones were isolated from a "HaeIII-AluI partial" soybean genomic library by screening with the nick-translated p3H9 insert. Two positive clones, designated 3H9.7.1. and 3H9.19.1, were identified after screening 700,000 recombinants. Restriction analysis (Fig. 5) indicated that the two clones contain overlapping regions of the same gene. The sizes of the two cloned genomic fragments were roughly 15 and 12 kb, respectively, for 3H9.7.1 and 3H9.19.1.

As seen in Fig. 5, EcoRI digestion of these clones generated a number of fragments within the insert. The location of the mRNA coding region was determined by hybridizing

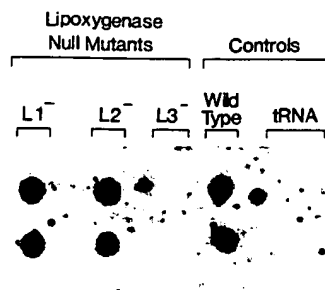


Fig. 4. Dot blots of RNAs from soybean lipoxigenase-null soybean mutants. Total polysomal RNAs from the indicated soybean lox phenotypes were dotted in duplicate (5  $\mu$ g per dot) onto Zetaprobe membrane. tRNA was applied in the same amount as a control for nonspecific binding. Hybridization was with p3H9/3', a gene-specific subclone of p3H9 containing 172 bp of 3' untranslated sequence (from positions 4975 to 5147 in the genomic sequence) in pGEM3. The plasmid was nick-translated and hybridized to the dot blot as described in the Materials and Methods

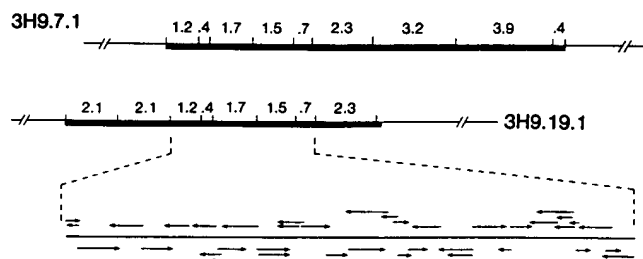


Fig. 5. *Eco*RI restriction map of cloned soybean genomic DNAs containing a lipoxigenase gene, and sequence determination strategy. Thick lines represent cloned genomic DNA. Thin lines are lambda DNA arms. Numbers located between short vertical lines represent the approximate distance, in kb, between adjacent *Eco*RI sites. The expanded map below clone 3H9.19.1 shows sequencing strategy. Arrows indicate direction and extent of analysis. Sequencing of upper or lower strand is shown by the positioning of the arrows above or below the solid line, respectively

individual nick-translated *Eco*RI fragments to Northern blots of soybean embryo RNA. The RNA-reactive probes from each clone were the 1.2, 0.4, 1.7, 1.5 and 0.7 kb fragments (data not shown). We subcloned each of these fragments from 3H9.19.1 into the *Eco*RI site of pBR322 and performed nucleotide sequence analysis according to Maxam and Gilbert (1980).

#### Structure of a lox-3 gene

We sequenced a 5,708 bp DNA segment encompassing all of the 1.2 kb *Eco*RI fragment and extending through the 0.7 kb *Eco*RI fragment (Figs. 5, 6). The structure of the lox gene was deduced following the additional cloning and sequencing of nearly full length cDNAs, and by performing primer extension and S1 protection experiments to determine the 5' boundary. The complete sequence and structure

of the gene is shown in Fig. 6. The results of the primer extension and S1 protection experiments shown in Fig. 7 are in good agreement, and support the assignment of the mRNA start site to the region around nucleotide 816 (see Fig. 6), although there is an apparent heterogeneity of several nucleotides in the start of the lox gene. This assignment is also supported by the relative position of the "TATA" box spanning nucleotides 781 through 786 (TATAAA). The initiation of transcription for most eukaryotic genes occurs some 30 nucleotides downstream from the TATA box. In addition, although a typical CAAT box is not present, a CAAG sequence is located roughly 80 nucleotides upstream (nucleotides 732–735) from the message start site. The sequence of 816 bp that extends upstream from the message start site was compared with upstream sequences of several other soybean genes that are expressed in embryos. Excluding the TATA or CAAT regions, no obvious homologies that may indicate functionally equivalent promoter or regulatory regions were evident.

The first ATG (methionine) codon after the start of transcription occurs at nucleotide 862 in the genomic sequence, indicating that the lox coding sequence is preceded by a 5' untranslated region of about 46 nucleotides. The nearly full-length cDNA that was sequenced, p3H9FL, extends 10 nucleotides into the 5' untranslated region and contains a 2571-nucleotide-long open reading frame. This is followed by a TGA stop codon and a 3' untranslated region of 148 nucleotides, which terminates at position 5101 in the genomic sequence. The smaller cDNA clone, p3H9, spans nucleotides 4676 through 5149 in the genomic sequence, demonstrating that at least two of the four consensus polyadenylation signals present in the 3' untranslated region at positions 4989, 5079, 5084 and 5119 are utilized. The 3' untranslated region of p3H9 is 196 nucleotides in length. It should be noted that both cDNA clones contain poly(A) tails. The open reading frame of p3H9FL encodes 857 amino acids, with an estimated molecular weight of 96,663 daltons, a size consistent with that observed after denaturing polyacrylamide gel electrophoresis of the in vitro translation product (Fig. 3).

A comparison of the genomic sequence with the cDNA sequence indicates that the gene consists of nine exons and eight introns. The exons range in size from 86 to 950 bp, while the intron sizes vary from 93 to 448 bp. The intron sequences bordering the intron-exon junctions all conform to the GT-AG rule. The GC content of the exons and introns is 41.6% and 28.3%, respectively.

#### Comparison of lox isozymes

A comparison of the deduced amino acid sequences of various lox isozymes derived from their corresponding cDNAs is shown in Fig. 8. The complete amino acid sequence for lox-1 is from Shibata et al. (1987) and the partial sequence for the isozyme designated "lox-1-like" is from Start et al. (1986). The designation "lox-1-like" was based on the fact that the clone encoding this sequence (pLX-65) showed reduced hybridization to RNA from a lox-1 null mutant,

Fig. 6. Nucleotide and predicted amino acid sequence of a soybean lipoxigenase-3 gene. Position 1 corresponds to the *Eco*RI site on the 5' side of the 1.2 kb *Eco*RI fragment of lambda clone 3H9.19.1. Major (\*, ●) and minor (+, ○) products from primer extension (\*, +) and S1 protection (●, ○) experiments are indicated. Single underline, CAAT-like box; double underline, TATA box; overline, consensus polyadenylation signals; arrows, locations of alternative poly(A) addition sites



.AATTGATTCTAATGTAAATTAATATCTAAAAATATATCTAAAAATCAATTATGAATTGAGAAATTAAT 70  
TTTAAACCTTAAACCAACATGCACTCAAACTTATTTCAACTTATAATTCATTTTAAAAATTTCTATAAC 140  
ATAAAATCAACATCTAAGAATATATTTAAAAATAGCTCAAAATCAATTTTAAACATCAACGAAACAC 210  
ACACTTATCAAACTATTATCTACGATTATCAATCAAACTTCAATTTTCATCGAAAAAAATATACU 280  
AAAGCTACCTATCTATCAAAACCACTTATAAGTTTATTTTAAATTAATTTTAAAGAAATTAATTA 350  
GTCAATAAAAATATTATTACATTAAGCAATCAAACTTACTGATTTTACTGCTAAAAAGCTCAAT 420  
AAACCTTATAAAAAATATCTGTATCTCTATATAATCTCTCTCTGCTTTAAACAAAGCTACTATTAA 490  
TAAAAAAATATTAATCTCTGAGTATATGAATCATGTTAAAAATTTATTATATAAAAGAGACTAAGAA 560  
AATGCTCTGGGAGATTTAAATTTATCAATTCACAAATGACTACTAACTACTACTACTAGTAAGA 630  
AGGAAATCTTGAATATAATTAATTTGATTTTATCTGCTTCTGCTTCTTGGCATTGCACTATAGCACT 700  
TTGCTTTCACATTTGAGCTGCTGCTGCTGCTCAACACATTTATGATCATCATCTTCTTCTTACTCTT 770  
CCTCTTAATTATAATACCAACAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840  
MetLeuGlyGlyLeuLeuHisArgGlyHisLysIleLysGlyThrValV 17  
TACTGTTGCT 910  
alLeuMetArgLysAsnValLeuAspValAsnSerValThrSerValGlyGlyIleIleGlyGlnGlyLe 40  
TGTTCATGCGCAACAATGCTTTCGATGCAATAGCGTAACACCGCTTGGGGAATTTATGCTCAAGGCT 980  
uAspLeuValGlySerThrLeuAspThrLeuThrAlaPheLeuGlyArgProValSerLeuGlnLeuIle 63  
CGACTTACCTTGGCTCAACCTCATACTTCTTACTGCTTCTTGGGCGACCGCTGCTCTCCAGCTTATT 1050  
SerAlaThrLysAlaAspA 70  
AGTCTACCAAGCTGATGCTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1120  
GCTTCAATTTCTACCTATCTTCTGCTCATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1190  
TTCAATTTCAATCTATCTCTCTCATATCTTCTGCTTCAACGGAGTTTGAATTCGCTGAAAAATATAACCA 1260  
TACACAGCTAGATGTTTATAATCATGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1330  
TTGATGTTTAAATTTATTAGTATACCATGCTGAGGATTTCTACTCTCTTACCAAGTATAGCTACGTTT 1400  
CACCGACCTCTATAATGTTTAAATTTTGTGTTATGTAAGTGAAGCACTAATCATGCTAGACGGCTT 1470  
laAsnGlyLysGlyLysLeuGly 77  
CATATTGCTTTTACTTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1540  
LysAlaThrPheLeuGlyGlyIleIleThrSerLeuProThrLeuGlyAlaGlyGlnSerAlaPheLysI 101  
AAGCTACCTTTTTCGAGGATCATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1610  
leAsnPheGluTrpAspAspGlySerGlyIleLeuGlyThrIleLysAsnPheMetGlnThrG 124  
TTAATTTTGAATGCGATCATGCGACTGCAATTTTGGACCATTTTATATCAAGAAATTTATGCAAACTG 1680  
uPhePheLeuValSerLeuThrLeuGluAspIleProAsnHisGlySerIleHisPheValCysAsnSer 147  
GTTTTCCTTCTGAGTTGACTCTTGAAGCATTTCAACCATGGAAGCATTCACCTTTGTTTGAATTCG 1750  
TrpIleTyrAsnAlaLysLeuPheLysSerAspArgIlePhePheAlaAsnGln 165  
TGCATTTTACAATGCAAACTCTTCAAAAGTACCGCATTTTCTTGGCAACAGGTCAAGTATAATTAAT 1820  
GAACTTAAATATACTTTTAAATCTTCAATACTTTGTTGTTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1890  
AAAAAGTTTTTGTCAATACCGGAGTGTCAATTAATGTGCTTATTTTGAAGTGTGTAACCTTAAAGCAT 1960  
ThrTyrLeuProSerGluThrProAlaProLeuVa 177  
CAATGAAAAATAAAGTAAACAAAAATATGCGAGCATATCTTCAAGTGCAGACACCGACTCCACTAGT 2030  
lLysTyrArgGluGluGluLeuHisAsnLeuArgGlyAspGlyThrGlyGluArgLysGluTrpGluArg 200  
CAATATAGAGAAGAGAGCTGCAATTAATGAAGGAGATGGAAGTGAAGAGCGCAAGAGCTGGGAAGC 2100  
ValTyrAspTyrAspValTyrAsnAspGlyLeuAspAspLysGlyGluAsnHisAlaArgProVal 224  
CTCTATGATTATGATCTCAATAGTATTAGCTGATGCGGATAAAGTGAAGAAATCATGCGCTGCTGCT 2170  
euGlyGlyAsnAspThrPheProTyrProArgArgGlyArgThrGlyArgLysProThrArgLysA 246  
TTGGAGGAAATGACACCTTTCTCTTATCTCTGAGGGAGCAACTGCTAGAAAAACCAAGGAAAGCTTA 2240  
GGCTGTATTTCTATTTTAAATTTACTCTATCAGACAGTCAAACTGCTTCTTCAATTTCTATAATATTAT 2310  
GAGCTCTAGCATAAATGACTTTTAAATATGCAAGAGTCATATACTTTGAATTCGTATATTGTTTAC 2380  
CTAGAAGACTTCTACATCTACATCTTAACATAAAGTTGACTGTTCTAATCAAGTAACTCAAAATATC 2450  
AAGGAGTACTTAGAAAAACCATTTTCAACTTTAAGCTGTTTCTTCAATATGAGTTTGAATTTCTG 2520  
spProAsnSerGluSerArgSerAsnAspVa 256  
TTTTTGTGCTAATAAGTTCTGCTATTTTCAATTAACAGATCTTAACTGAGACTAGGACCAATGATGT 2590  
lTyrLeuProArgAspGluAlaPheGlyHisLeuLysSerSerAspPheLeuThrTyrGlyLeuLysSer 279  
TTATCTTCAAGAGATGAGGCTTTTGGACACTTGAAGTCACTGCACTTCTTACTTATGCACTAAATCC 2660  
ValSerGlnAsnValLeuProLeuGlnSerAlaPheAspLeuAsnPheThrProArgGluPheAspA 303  
GTATCTCAAAATGCTTCTCCATTATGCAATTCGCTTTGATTTGAATTTACACACCGCTGAGTTTGATA 2730  
rPheAspGluValHisGlyLeuTyrSerGlyGlyIleLysLeuProThrAspIleIleSerLysIleSe 326  
GCTTTGATGAAGTTCATGAGCTATTTTACGGGGAATTAAGCTGCGCAAGATATAATCAGCAAGATTAC 2800  
rProLeuProValLeuLysGluIlePheArgThrAspGlyGluGlnAlaLeuLysPheProProProLys 349  
TCCACTACCGCTGCTTAAAGAACTCTTCCGAACTGATGCTGCAAGCGGCTTAAGTTTCTCTCTCTAA 2870  
ValIleGlnV 353  
GTAATTCAGCTATCTCATATAATTAAGTACCTGCTATCTAAAAATGCGGCTAGCAATTAACAGCATGGACT 2940  
alSerLys 355  
GCTAATAATGATATCTGAATTAATCTATTTCTACTAAGCATTTTGTATTTATTTGAGGAGTGAAG 3010  
SerAlaTrpMetThrAspGluGluPheAlaArgGluMetLeuAlaGlyValAsnProAsnLeuIleArgC 379  
TCTGATGATGACTGATCAAGAAATTTGCAAGAGAAATGCTTGGTGTAAATCCAACTGATTCGTT 3080  
ysLeuLys 381  
GCTTAAAGCTTGAATTTTCTTGGATGCTCAACTTTCTGATGATATCTTGAATGGCAGGATGCCCAAT 3150  
GluPheProProArgSerLysLeuAspS 391  
CGAGGAGCATATATATTCTTATAATTTATTTATGAGGAGTTCCTCCAGCAAGCACTAGATA 3220

erGlnValTyrGlyAspHisThrSerGlnIleThrLysGluHisLeuGluProAsnLeuGluGlyLeuTh 414  
GCCAAGCTCTATGCTGATCACTACTCAAAATACCAAGAACACCTAGAGGCCCACTAGACGGCTCAC 3290  
rValAspPhe 417  
TGAGATGAGGTATATAATATGATAAACTTGTGTTTCAAAATAAATTTTATTTGATGAATTTCTGTTATT 3360  
TTATGCTGCTAGTTTTATTTTATGATGATGTTTCTATGAATAAAGCAATAATGCTATTGCTTGGTAA 3430  
AlaIleGlnAsnLysArgLeuPheLeuGlyHisHisAspProIleMetProTyrLeuArgArgIle 440  
GCCAATTCAAAACAGAGATGTTCTTACTACGTCATCATGACCAATCTGCGCAATTTGAGGCGAATA 3500  
AsnAlaThrSerThrLysAlaTyrAlaThrArgThrIleLeuPheLeuLysAsnAspGlyThrLeuArgP 464  
AATGCAACCTCCACAAAGGCTTATGCTACGAGAACCTGCTTTTCTGAAAAATCAAGCACTTTAAGAC 3570  
roLeuAlaIleGluLeuSerLeuProHisProGlnGlyAspGlnSerGlyAlaPheSerGlnValPheLe 487  
CACTTGCCATAGAGTTGAGTTTGGCAGATCTCAGGAGATCAATCTGCTGCTTTTACTCAAGCTTTTCT 3640  
uProAlaAspGluGlyValGluSerSerIleTrpLeuLeuAlaLysAlaTyrValValValAsnAspSer 510  
GCTCGAGATGAAGCTGTTGAAGTCTTATTTGCTGCTAGCAAGCTTATGCTACTTGTGAATGACTCG 3710  
CysTyrHisGlnLeuValSerHisTr 519  
TCTATCATCAACTTCTGAGCCATCGGTATACATTTCTCAAGAAATATAAAAAATAAACTTAA 3780  
pLeuAsnThrHisAlaValVa 526  
TATATATATCTAGAAAAATGATTATTAATATCTTTGATGCGATAAGCTTAAACACTCATGCACTGCT 3850  
lGluProPheIleIleAlaThrAsnArgLysLeuSerValValHisProIleTyrLysLeuHisPro 549  
TGAGCCATTCATCATAGCAACAAACAGGCTCTGCTGTTTCCACCTATTTTAACTCTCTTACCCT 3920  
HisTyrArgAspThrMetAsnIleAsnGlyLeuAlaArgLeuSerLeuValAsnAspGlyGlyValIleC 573  
CACTATCTGACACCATGACATATAATGCGCTTCTGCTGTTACTGCTGCAACGAGCTGGCTTATAG 3990  
luGlnThrPheLeuTrpGlyArgTyrSerValGluMetSerAlaValValTyrLysAspTrpValPheTh 596  
AACAAATATTTTGGGGAGGATTTCTGCGAAATGCTGCTGATGTTTACAGGATTGGGTTTTAC 4060  
rAspGlnAlaLeuProAlaAspLeuLysAr 607  
AGATGAAGCTTGCCTGCTGACTTATAAAAGCTAAATCATATATCTTACAATAGCAAAACATGAACTC 4130  
AATTTAATTTATAGTTAAACAAAAATGAACACACTTTGATGTTATATAATTAATGTCATGATATAG 4200  
gGlyMetAlaIleGluAspProSerCysProHisGlyIleArgLeuValIleGluAspTyrProTyrAla 630  
AGCAATGCAATTCAGGATCATCTGCTGCTCATGCGCTTCTGATAGAGGACTACCTTATGCT 4270  
ValAspGlyLeuGluIleTrpAspAlaIleLysThrTrpValHisGlyTyrValPheLeuTyrTyrLysS 654  
GTTGATGCACTTCAGATATGCGATCTATCAAGAGATGCGTCAATACCTTTTCTTCTACTACAAAT 4340  
erAspAspThrLeuArgGluAspProGluLeuGlnAlaCysTrpLysGluLeuValGluValGlyHisG 677  
CAGATGACACACTTAGAGAGATCTGCAAGCTGCAAGCTGCTGCAAGCAACTCTGAGGCTGGCTCATGC 4410  
yAspLysLysAsnGluProTrpTrpProLysMetGlnThrArgGluGluLeuValGluAlaCysAlaIle 700  
AGCAAGAAAAATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4480  
IleIleTrpThrAlaSerAlaLeuHisAlaAlaValAsnPheGlyGlnTyrProTyrGlyGlyLeuIleL 724  
ATCATATGAGCTGCTGAGCACTTATGCACTGCTTAAATTTGCACTATCCCTATGAGCTTTTAACT 4550  
euAsnArgProThrLeuSerArgArgPheMetProGluLysGlySerAlaGluTyrGluGluLeuArgLy 747  
TAAACGCTCAACTCTTACTAGCGATTCATGCTGAGAAAGCTTCTGCTGACTATGAGGAGCTGAGGAA 4620  
sAsnProGlnLysAlaTyrLeuLysThrIleThrProLysPheGlnThrLeuIleAspLeuSerValIle 770  
GAATCCCAAGAGGCTTACTTGAAGACTATTACCAAGTCTTCAAGCTTATTGACCTTTCTGTTATA 4690  
GluIleLeuSerArgHisAlaSerAspGluValTyrLeuGlyGluArgAspAsnProAsnTrpThrSerA 794  
GAAATCTTCTCAAGCATTCATCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 4760  
spThrArgAlaLeuGluAlaPheLysArgPheGlyAsnLysLeuAlaGlnIleGluAsnLysLeuSerG 817  
ATACAGAGCGTTAGCGCTTTTAAAGCTTTGAAATAACTGGCAAAATGAGGAATAAACTCTCAGA 4830  
uArgAsnAspGluLysLeuArgAsnArgCysGlyProValGlnMetProTyrThrLeuLeuLeuPro 840  
GAGAAACACGATGAGAACTGAGAAACGCTTGTGAGGAGCTTCAATGCTTATCTGCTGCTTTCGCT 4900  
SerSerLysGluGlyLeuThrPheArgGlyIleProAsnSerIleSerIleEnd 857  
TCTACTAGGAAGGATTAATCTTCAAGCAATTCCTCAAGCATGCTCTATCTGAGGAGCTTGTGCTTCC 4970  
TTAAGTTTGTGCTTTCATAAAGATAGAGAGAGCAATCTGAAATTTCTCATCAACCTGCAGATC 5040  
GTAATCTAGCTGTTTGAAGCTTTCATGTTTGTGCTTTTAAATAAAAGAGCTTGTTTGTAATAATGT 5110  
TCAGATCAATAAAGCATTTACTGCTTGTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5180  
AAATAAGCTTGGCAAAAGCTAGAACGGCTGATAACCCATTTGTTCAATATAATCGCTAATCTTTGT 5250  
TTTATTTTGTTCACAGAGATATAAGAAACAAATATATAATGCAAACTTTATTTAAAAATATATTC 5320  
GCTAAGAGCTTGTGCTGCTGAGCAAGATTTGATTTTCAATCAAAACGGTCAACCTATTAATAAAT 5390  
CACAGGAGCACTAAGTACCAAAATATGCTTATTAACATATAAATGAACTTTCAAAATGAAGACA 5460  
ACCTGCTGAGGAGGATACTTTTATCATTCATGCTTTTCCGCTTCTTCTGATGATATATCTTGC 5530  
AAATCAAACTCTGTTTCACTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 5600  
TATAAAAAAGAAATCTTGAAGCGCTTAACTCGAACCAACCTACTCAATAGAAATTTGAAGCTTT 5670  
GGGCAATGCTTGAATTTGCAATACATCTTGAATGCTG 5709



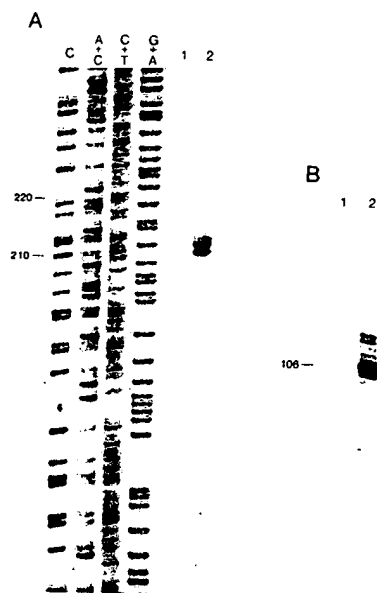


Fig. 7. Identification of the 5' end of the lipoxygenase gene by primer extension and S1 protection studies. Labeled DNA fragments used as hybridization probes were annealed in the presence of yeast tRNA (lanes 1A and 1B) or total polysomal RNA (lanes 2A and 2B), and either extended with reverse transcriptase (A) or treated with nuclease S1 and the products were analyzed by acrylamide gel electrophoresis (B). The autoradiogram in (A) shows the chemical degradation products used as size markers, with base-specific cleavages indicated above the appropriate lane. Numbers indicate the size of fragments in nucleotides

while no reduction was seen in hybridizations with RNAs from other *lox* nulls. However, B. Axelrod has recently isolated a cDNA clone that has the same sequence as pLX-65, and identified it as encoding a *lox*-2 isozyme; this was based on a comparison of the deduced amino acid sequence with extensive amino acid sequence data obtained from purified *lox*-2 protein (personal communication). The difference between the two assignments of *lox* isozymal type may possibly be a consequence of the high degree of homology between the *lox*-1 and *lox*-2 sequences. The deduced partial amino acid sequence of the "lox-1 like" (*lox*-2) protein shows 83% homology with the corresponding sequence of *lox*-1. At the nucleotide level, the coding regions are 89% homologous. Cross hybridization of the pLX-65 probe (which contained some coding region) with *lox*-1 mRNA in the *lox*-2 null mutant could have masked the absence of the mRNA for *lox*-2, particularly so if *lox*-1 mRNA levels are substantially higher than those of *lox*-2 in soybean embryos. We are currently planning to address the issue of relative *lox* mRNA levels in soybean embryos through the use of gene-specific oligonucleotide primers.

Additional confirmation that the gene in 3H9.19.1 encodes a *lox*-3 isozyme comes from a comparison of the deduced *lox* amino acid sequences with the determined amino acid sequence data obtained by microsequencing of purified *lox*-3 peptides (data provided by B. Axelrod). As can be seen in Fig. 8, the deduced *lox*-3 amino acid sequence shows 100% homology with the actual amino acid sequences of four different regions of *lox*-3. On the other hand, regions of dissimilarity with the *lox*-1 and "lox-1-like" (*lox*-2) sequences are evident. This finding is consis-

tent with data presented earlier, showing that the gene specific clone p3H9/3' does not hybridize to RNA from a *lox*-3 null mutant (Fig. 4). Thus, we are certain we have a gene encoding a *lox*-3 isozyme, although we are unable to determine whether we have a *lox*-3A or *lox*-3B isozyme from the available data.

As shown in Fig. 8, the deduced protein sequences of *lox*-3 and *lox*-1 show 70% homology, while the *lox*-3 and "lox-1-like" (*lox*-2) deduced sequences have 73% homology. The corresponding nucleotide sequence homologies over the same regions are 74% and 73%, respectively. The sequence of *lox*-1 indicates that it consists of 838 amino acids while the *lox*-3 sequence encodes 857 amino acids. The length difference of 19 amino acids is due, principally, to the presence of codons for 17 additional amino acids within the first exon of *lox*-3. The amino acid sequence within exon 6 of *lox*-3 (amino acids 382–417) also shows a high degree of divergence from that of *lox*-1. Exons 1 and 6 of *lox*-3 show roughly 44% and 42% nonhomology, respectively, with their corresponding regions in the *lox*-1 sequence, while the remaining exons differ only by 16%–25%.

We tried to identify putative functional domains on the basis of exon-intron structure, hoping that this might provide some insight into the enzymatic characteristics of these isozymes, but with the exception of the differences noted above, we had no success. For instance, hydropathy analysis performed on *lox*-3 revealed no obvious hydrophobic lipid binding site(s) in the above locations, although the region spanning amino acids from approximately 500 through 700 has a slightly more hydrophobic nature than the remainder of the protein (data not shown). This latter region corresponds to exon 8 and a small portion of exon 9 of *lox*-3.

As mentioned earlier, *lox*-2 and -3 are generally classified together as type 2 enzymes, while *lox*-1 is placed in a separate class, designated type 1. The biochemical and enzymatic properties of *lox*-1 differ markedly from those of the other two *lox* isozymes. It is interesting, then, that *lox*-1 and *lox*-2 have a substantially higher degree of homology than do *lox*-2 and *lox*-3. Moreover, a comparison of the 3' untranslated regions of cDNA clones for these three isozymes reveals that *lox*-1 and -2 are more closely related to each other than either one is to *lox*-3 (Fig. 9). Although the nucleotide sequences of the coding regions of the three *lox* isozymes show a minimum of 70% homology, the 3' untranslated regions of *lox*-1 and *lox*-2 show only 47% and 49% homology, respectively, with *lox*-3. On the other hand, a comparison of this region between *lox*-1 and -2 shows 76% homology. Therefore, the 3'-end of *lox*-3 gene is substantially more divergent from the corresponding region of the cDNA sequences for *lox*-1 and -2. In addition, the *lox*-1 and -2 genes not only have a higher degree of structural similarity, but they are also genetically linked. Soybean *lox*-1 and -3 segregate independently, but the genes for *lox*-1 and -2 do not (Hildebrand and Hymowitz 1981; quoted in Casey et al. 1985). Thus, although the designation "lox-1-like" is a misnomer in a formal sense, it is a curiously appropriate designation when viewed in light of the above considerations.

We hope a complete characterization of the entire gene family will make it possible to address the means by which specific lipoxygenase genes are regulated, as well as to define how the different isozymes influence the overall physiology of the plant.

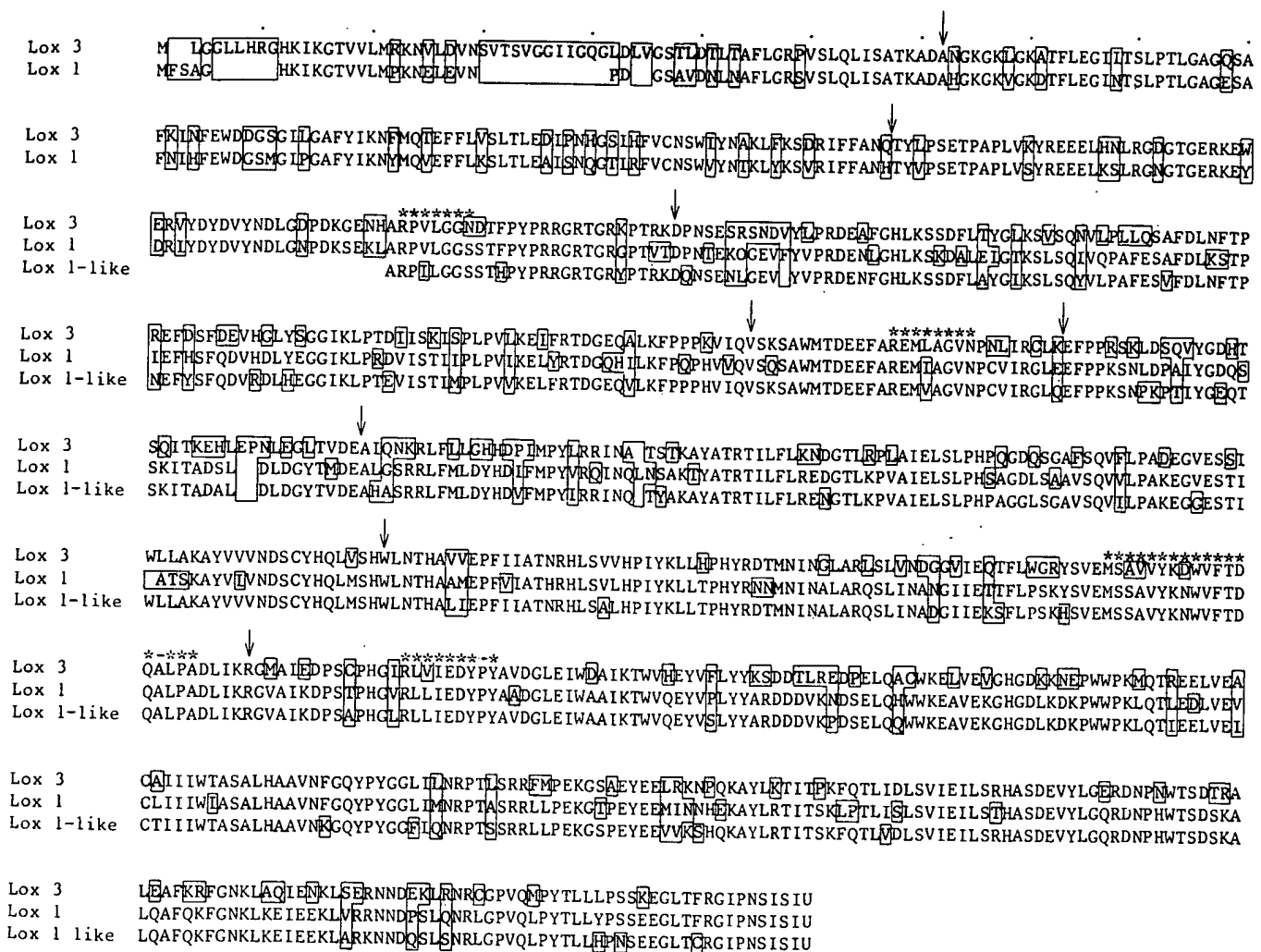


Fig. 8. Comparison of the deduced amino acid sequences for lipoxygenase isozymes. Sequences are aligned for maximum homology. Shaded areas indicate differences. Arrows indicate borders between adjacent exons. Residues determined by microsequencing of lox-3 peptides are indicated by asterisks. Residues not determined by microsequencing are shown by dashes (-). Determined and deduced lox-3 sequences were 100% homologous. The lox-1 sequence is from Shibata et al. (1987). "Lox-1-like" is the designation given by Start et al. (1986) to the clone pLX-65 and includes corrections of their earlier published sequence data (D. Hildebrand, personal communication). As discussed in the text, this latter sequence is most probably that of the lox-2 isozyme.

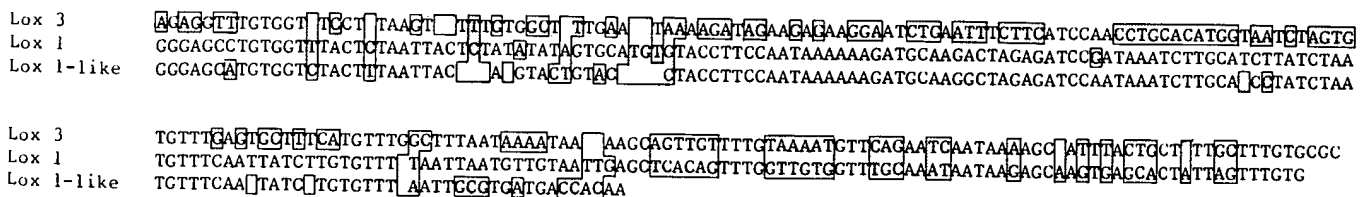


Fig. 9. Comparison of 3' untranslated regions of mRNAs encoding different lipoxygenase isozymes. The first nucleotide shown follows the stop codon of each isozyme. The "lox-1-like" sequence has been corrected as described in Fig. 8. Shaded areas reflect differences between the sequences.

**Acknowledgements.** We wish to acknowledge and thank Dr. Robert Goldberg for supplying the mid-maturation stage soybean embryos and the phage genomic library. Dr. T. Kaye Peterman supplied the affinity purified lipoxygenase antibody. Dr. David Hildebrand provided the lipoxygenase, null mutants of soybean embryos. Dr. Bernard Axelrod supplied purified lipoxygenase protein, as well as the determined amino acid sequence data for lox-3 protein. Drs. Hildebrand and Axelrod were most generous with information. We thank Drs. John Grula and David Anderson for carefully read-

ing the manuscript and Mrs. Mary Strother for its typing. This work was supported by the J.G. Boswell Company (Los Angeles).

## References

- Axelrod B (1974) Lipoxygenases. *Adv Chem Ser* 136:324-348
- Axelrod B, Cheesebrough TM, Laakso S (1981) Lipoxygenase from soybean. *Methods Enzymol* 71:441-451

- Casey R, Domoney C, Nielsen NC (1985) Isolation of a cDNA clone for pea (*Pisum sativum*) seed lipoxygenase. *Biochem J* 232:79-85
- Christopher J, Axelrod A (1971) On the different positioned specificities of peroxidation of linoleate shown by two isozymes of soybean lipoxygenase. *Biochim Biophys Res Commun* 44:731-736
- Christopher J, Pistorius E, Axelrod B (1970) Isolation of an isozyme of soybean lipoxygenase. *Biochim Biophys Acta* 198:12-19
- Christopher JP, Pistorius EK, Axelrod B (1972) Isolation of a third isoenzyme of soybean lipoxygenase. *Biochim Biophys Acta* 284:54-62
- Eskin NAM, Grossman S, Pinsky A (1977) Biochemistry of lipoxygenase in relation to food quality. *Crit Rev Food Sci Nutr* 9:1-40
- Galliard T, Chan HWS (1980) Lipoxygenases. In: Conn EE, Stumpf PK (eds) *The Biochemistry of Plants*, vol 4. Academic Press, New York, pp 131-161
- Goldberg RB, Hoschek G, Tam SH, Ditta GS, Breidenbach RW (1981) Abundance, diversity, and regulation of mRNA sequence sets in soybean embryogenesis. *Dev Biol* 83:201-217
- Grunstein M, Hogness DS (1975) Colony hybridization: A method for the isolation of cloned DNAs that contain a specific gene. *Proc Natl Acad Sci USA* 72:3961-3965
- Gubler U, Hoffman BJ (1983) A simple and very efficient method for generating cDNA libraries. *Gene* 25:263-269
- Hildebrand DF, Hymowitz T (1981) Two soybean genotypes lacking lipoxygenase-1. *J Am Oil Chem Soc* 58:583-586
- Hildebrand DF, Kito M (1984) Role of lipoxygenases in soybean seed protein quality. *J Agric Food Chem* 32:815-819
- Hildebrand DF, Rodriguez JG, Brown GC, Luu KT, Volden CS (1986) Peroxidative responses of leaves in two soybean genotypes injured by twospotted spider mites (*Acari: Tetranychidae*). *J Econ Entomol* 79:1459-1465
- Kanofsky JR, Axelrod B (1986) Singlet oxygen production by soybean lipoxygenase isozymes. *J Biol Chem* 261:1099-1104
- Leshem YY (1984) Interaction of cytokinins with lipid-associated oxy free radicals during senescence: a prospective mode of cytokinin action. *Can J Bot* 62:2943-2949
- Lupu R, Grossman S, Cohen Y (1980) The involvement of lipoxygenase and antioxidants in pathogenesis of powdery mildew on tobacco plants. *Physiol Plant Pathol* 16:241-248
- Maniatis T, Fritsch EF, Sambrook J (1982) *Molecular cloning, a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
- Maxam AM, Gilbert W (1980) Sequencing end-labeled DNA with base specific chemical cleavages. *Methods Enzymol* 65:499-560
- Mustakas LC, Albrecht WJ, McGlee JE, Black LT, Bookwalter GN, Griffin JJ (1969) Lipoxidase deactivation to improve stability, odor and flavor of full-fat soy flours. *J Am Oil Chem Soc* 46:623-626
- Needleman P, Turk J, Jakschik BA, Morrison AR, Lefkowitz JB (1986) Arachidonic acid metabolism. *Annu Rev Biochem* 55:69-102
- Ocampo CA, Moerschbacher B, Grambow HJ (1986) Increased lipoxygenase activity is involved in the hypersensitive response of wheat leaf cells infected with avirulent rust fungi or treated with fungal elicitor. *Z Naturforsch* 41c:559-563
- Samuelsson B (1983) Leukotrienes: Mediators of immediate hypersensitivity reactions and inflammation. *Science* 220:568-575
- Sanger F, Micklen S, Coulson AR (1977) DNA sequencing with chain terminating inhibitors. *Proc Natl Acad Sci USA* 74:5463-5467
- Shibata D, Stęczko J, Dixon JE, Hermodson M, Yazdanparast R, Axelrod B (1987) Primary structure of soybean lipoxygenase-1. *J Biol Chem* 262:10080-10085
- Start WG, Ma Y, Polacco JC, Hildebrand DF, Freyer GA, Altshuler M (1986) Two soybean seed lipoxygenase nulls accumulate reduced levels of lipoxygenase transcripts. *Plant Mol Biol* 7:11-23
- Theorell H, Holman RT, Akeson A (1947) Crystalline soybean lipoxidase. *Arch Biochem* 14:250-252
- Veldink GA, Vliegthart JFG, Boldingh J (1977) Plant lipoxygenases. *Prog Chem Fats Other Lipids* 15:131-166
- Vernooy-Gerritsen M, Bos ALM, Veldink GA, Vliegthart JFG (1983) Localization of lipoxygenases -1 and -2 in germinating soybean seeds by an indirect immunofluorescence technique. *Plant Physiol* 73:262-267
- Wolf WJ (1975) Lipoxygenase and flavor of soybean protein products. *J Agric Food Chem* 23:136-141
- Yenofsky R, Bergmann I, Brawerman G (1982) Cloned complementary deoxyribonucleic acid probes for untranslated messenger ribonucleic acid components of mouse sarcoma ascites cells. *Biochemistry* 21:3909-3913

Communicated by R.B. Goldberg

Received October 22, 1987

- (1986); R. Heikkilä *et al.*, *Nature* 328, 445 (1987); J. T. Holt, R. L. Redner, A. W. Nienhuis, *Mol. Cell. Biol.* 8, 963 (1988); E. L. Wickstrom *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 85, 1028 (1988).
7. A. M. Gewirtz and B. Calabretta, *Science* 242, 1303 (1988).
  8. D. Caracciolo, D. Venturelli, M. Valtieri, C. Peschle, A. M. Gewirtz, B. Calabretta, in preparation.
  9. Cultures were supplemented with recombinant erythropoietin [rh Epo (3 U/ml)] for the CFU-E assay, and Epo (3 U/ml), GM-CSF (5 ng/ml), and IL-3 (20 U/ml) for BFU-E. CFU-GM progenitors were assayed in 0.8% methylcellulose in the presence of GM-CSF (10 ng/ml) and IL-3 (20 U/ml); CFU-G progenitors were assayed in 0.3% agar in the presence of 10% conditioned medium of G-CSF-producing Chinese hamster ovary cells [D. J. Tweardy *et al.*, *Oncogene Res.* 1, 209 (1987)]. CFU-E and CFU-G colonies were analyzed after 7 and 9 days growth, respectively. CFU-GM and BFU-E colonies were analyzed after 14 days of culture.
  10. E. Shivelman, B. Lifshitz, R. P. Gale, A. Roe, E. Canaani, *Cell* 47, 277 (1986).
  11. A. Bernards, C. M. Rubin, C. A. Westbrook, M. Paskind, D. Baltimore, *Mol. Cell. Biol.* 7, 3231 (1987).
  12. I. K. Hariharan and J. M. Adams, *EMBO J.* 6, 115 (1987).
  13. The sequence of *c-abl 1* (sense) is 5'-GGGCAGCA-GCCTGAAAA-3'; the sequence of *c-abl 2* (antisense) is 5'-TTTTCAGGCTGCTGCCC-3'; the sequence of *c-abl 3* (sense) is 5'-TTGAGATCTG-CCTGAAG-3'; the sequence of *c-abl 4* (antisense) is 5'-CTTCAGGCAGATCTCAA-3'; the sequence of *c-abl 5* (sense) is 5'-GCCTTCAGCGGCCAGTA-3'; and the sequence of *c-abl 6* (antisense) is 5'-TA-CTGCGCGCTGAAGGGC-3'. The sequence of *bcrl 1* (sense) is 5'-GTGGACCCGGTGGGGTTC-3', and the sequence of *bcrl 2* (antisense) is 5'-GAAGC-CCACCGGGTCCAC-3'. Analysis of the homology between the synthesized oligomer and the primate sequences that are present in the GenBank database revealed that each synthetic oligomer could only be fully complementary to its own specific mRNA (the highest homology found was 14%).
  14. C. I. Civin *et al.*, *J. Immunol.* 133.
  15. J. Brandt, N. Baird, L. Lu, E. Sr. *J. Clin. Invest.* 82, 1017 (1988).
  16. D. Ferrero *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 80, 4813 (1983).
  17. M. Valtieri, A. M. Gewirtz, B. Calabretta, unpublished results.
  18. R. K. Saiki *et al.*, *Science* 239, 487 (1988).
  19. S. J. Collins and M. T. Groudine, *Proc. Natl. Acad. Sci. U.S.A.* 80, 4813 (1983).
  20. W. M. Wormington, *ibid.* 83, 8639 (1986); P. Dash, I. Lotan, M. Knapp, E. R. Kandel, P. Goel, *ibid.* 84, 7896 (1987).
  21. D. A. Rappolee *et al.*, *Science* 241, 708 (1988).
  22. S. V. Suggs, R. B. Wallace, T. Hirose, E. H. Kawashima, K. Itakura, *Proc. Natl. Acad. Sci. U.S.A.* 78, 6613 (1981).
  23. We thank R. Baserga and C. M. Croce for critical reading of the manuscript, J. K. deRisi for oligodeoxynucleotides synthesis, and R. Aringhaus for his gift of the antiserum to a *c-abl* peptide. Supported by grants CA46782, CA36896, CA01324 from the National Cancer Institute and grant CH-455 from the American Cancer Society. A.M.G. is a

## Claim 63 Ecdysteroid UDP- glycosyl transferase

# A Baculovirus Blocks Insect Molting by Producing Ecdysteroid UDP-Glucosyl Transferase

DAVID R. O'REILLY AND LOIS K. MILLER

The predicted amino acid sequence of a newly identified gene of the insect baculovirus *Autographa californica* nuclear polyhedrosis virus was similar to several uridine 5'-diphosphate (UDP)-glucuronosyl transferases and at least one UDP-glucosyl transferase. Genetic and biochemical studies confirmed that this gene encodes an ecdysteroid UDP-glucosyl transferase (egt). This enzyme catalyzes the transfer of glucose from UDP-glucose to ecdysteroids, which are insect molting hormones. Expression of the *egt* gene allowed the virus to interfere with normal insect development so that molting was blocked in infected larvae of fall armyworm (*Spodoptera frugiperda*).

BACULOVIRUSES CONSTITUTE A LARGE group of DNA-containing viruses that infect only invertebrate hosts. These viruses, many of which infect pest lepidopteran species, are of particular interest because of their potential as biological control agents (1). Studies of *Autographa californica* nuclear polyhedrosis virus (AcMNPV), the model system for baculovirus research, are revealing the molecular mechanisms by which the virus implements its replication strategy. Most of the genes identified to date are involved in the interaction of the virus with its host cell, but little is known of the molecular aspects of infection at the organismal level. We now report that AcMNPV has a gene that allows the virus to manipulate the hormonal regulation of development of its larval host.

The region of the AcMNPV genome containing the *egt* gene, spanning 8.4 to 9.6 map units, first came to our attention as a

hypermutable region in serially propagated viruses (2). The sequence of this region revealed an open reading frame that could encode a 57-kD polypeptide of 506 amino acids (Fig. 1). The *egt* product shares 21 to 22% amino acid sequence identity with several mammalian UDP-glucuronosyl transferases, also shown in Fig. 1. In mammals, the UDP-glucuronosyl transferases catalyze the transfer of glucuronic acid to a wide variety of exogenous and endogenous lipophilic substrates (3). This conjugation reaction is of critical importance in the detoxification and safe elimination of a multitude of drugs and carcinogens. In addition, the normal metabolism and disposal of various endogenous compounds, such as bilirubin and steroid hormones, proceed through their conjugation with glucuronic acid. Available evidence on insect systems indicates that sugar conjugation reactions of this type involve glucose rather than glucuronic acid transfer (4). No sequences of UDP-glucosyl transferase genes were available in GenBank at the time of our search, but the

**Table 1.** Substrate specificity of *egt* gene product. Substrates were incubated in the presence of medium derived from appropriately infected cells and 0.05  $\mu$ Ci UDP-[U- $^{14}$ C]glucose (312.5 mCi/mmol) for 1 hour at 37°C. All substrates were used at a concentration of 1 mM. Other conditions were as described in the legend to Fig. 2. Amounts of glucose transferred were calculated after scintillation counting of the appropriate regions of the chromatography plates. A hyphen indicates that less than 1% of the glucose was transferred.

Substrate	wt (pmol glucose trans- ferred)	vEGTZ (pmol glucose trans- ferred)
p-Aminobenzoic acid	—	—
Bilirubin	—	—
Chloramphenicol	—	—
Cholesterol	—	—
Diethylstilbestrol	—	—
Ecdysone	155.7	—
$\beta$ -Estradiol	—	—
20-Hydroxyecdysone	87.8	—
8-Hydroxyquinoline	—	—
Makisterone A	89.6	—
(-)-Menthol	—	—
Methylumbelliferol	—	—
$\alpha$ -Naphthol	—	—
p-Nitrophenol	—	—
Phenolphthalein	—	—
Testosterone	—	—
$\alpha$ -Tetralol	—	—

sequence of a UDP-glucosyl transferase gene from *Zea mays* (maize) was subsequently reported (5). The COOH-terminal portion of this protein also displays homology to *egt* and to mammalian UDP-glucuronosyl transferases (Fig. 1).

Mammalian UDP-glucuronosyl transferases are known to be membrane-bound, and the amino acid sequences of these proteins

contain both a putative signal sequence at the NH<sub>2</sub>-terminus and a halt-transfer sequence at the COOH-terminus (6-8). Although the predicted amino acid sequence of *egt* contains a putative signal sequence, the protein is approximately 30 amino acids shorter than the mammalian enzymes at the COOH-terminus (Fig. 1) and lacks the residues that constitute the halt-transfer sequence. These facts suggested the possibility that the *egt* protein is secreted.

The predicted amino acid sequence of the AcMNPV *egt* protein suggested to us that the enzyme might conjugate and eliminate some compound in the insect hemolymph. As in mammals, a wide variety of both exogenous and endogenous substrates are prone to conjugation in insect systems (4). Recently, an ecdysteroid-glucose conjugate was reported (9). Therefore, several ecdysteroids were included among the compounds tested as potential substrates for *egt*-mediated conjugation (Fig. 2).

*Spodoptera frugiperda* (fall armyworm) cells, which serve as hosts for AcMNPV, were infected with wild-type (wt) AcMNPV L-1

(10) or the recombinant virus vEGTZ. This mutant virus differs from wt AcMNPV only in the disruption of the *egt* gene by the insertion of the *Escherichia coli*  $\beta$ -galactosidase gene (Fig. 1). Twelve hours after infection, lysates or media were incubated in the presence of [<sup>3</sup>H]ecdysone and either UDP-glucose (Fig. 2) or UDP-glucuronic acid. Samples were then analyzed for ecdysone derivatives with altered polarities by thin-layer chromatography. A novel ecdysone derivative (EG) was observed only when the lysate or extracellular medium of wt AcMNPV-infected cells was used (Fig. 2A). When UDP-[U-<sup>14</sup>C]glucose was included in the reaction mix, it was possible to detect both <sup>3</sup>H and <sup>14</sup>C in the product (Fig. 2B), confirming that the conjugation reaction involves the transfer of glucose from UDP-glucose to ecdysone. The ecdysteroid UDP-glucosyl transferase activity was found predominantly in the extracellular medium. No ecdysteroid-conjugating activity was produced by mock-infected or vEGTZ-infected cells.

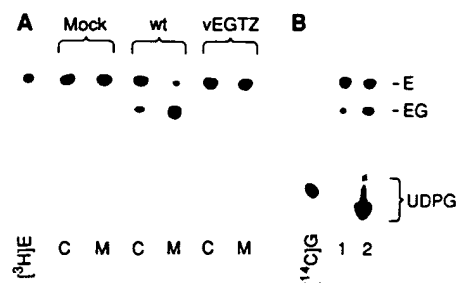
Although several ecdysteroids could function as substrates for this enzyme, none of

the other compounds tested were conjugated to a significant extent (Table 1). The ecdysteroids ecdysone, 20-hydroxyecdysone, and makisterone A were conjugated only in the presence of medium from wt-infected cells when UDP-glucose was included in the reaction mix. They were not conjugated when UDP-glucuronic acid was used (11) or when they were incubated in the presence of medium from vEGTZ (Table 1) or mock-infected cells (11). Thus, we conclude that the product of the *egt* gene is an ecdysteroid UDP-glucosyl transferase that is secreted into the extracellular medium by wt virus-infected cells.

Hemolymph titers of ecdysteroids fluctuate in a cyclic fashion to regulate both larval and larval-pupal molts (12). Since glucose conjugation is suspected to act as an

**Fig. 1.** Similarity of *egt* gene product to several UDP-glucuronosyl transferases and a UDP-glucosyl transferase. The predicted amino acid sequence of *egt* (E) (13) was compared to human (H) (6), mouse (M) (7), and rat (R) (8) UDP-glucuronosyl transferases by the use of the FASTP algorithm (14) as implemented by International Biotechnologies, Inc. The homology of a plant UDP-glucosyl transferase (Z) (5) to the above proteins is also displayed. Uppercase letters denote exact matches; lowercase letters correspond to substitutions that occur frequently among related proteins (15). Substitutions that occur infrequently are indicated by a dot. Gaps in a sequence are designated by hyphens; a caret marks where an amino acid has been deleted from the sequence. The amino acid sequence between the vertical arrows was replaced by the  $\beta$ -galactosidase gene in vEGTZ.

E	MTILCWLALLS-----TLTAVNAANILAVFPTPAYSHHIVKYVIEALAEKCHNVTV	
H	Msm....ALL.....fss.s.gkvL-V.PT-.fsh.m..K..ld.L.qr.HeVTV	
M	M.....ALL.....f.svk.gkvL-V.P-.fsh.m..K..ld.L.qr.HeVTV	
R	M.....ALF.....f.s.h.gkvL-V.P-.fsh.m..K..ld.L.qr.HeVTV	
E	VKP-KLFAYSTKTYCGNITEI-NADMSVE-----QYKKLVANSAMFRKRGVSDTDT	
H	l.s....isf..ns.....Ev..lt.....KqLV..A....kd.....s	
M	lrP...y...K...G...E...t.v.s.d.....K.v...t....Rd.....	
R	lKP...F...K...d...Ei.st.is.d.....K.L...t....Rd.....	
E	VTAANYLGLIEMFKDQF-DNINVRNLIANNQ---TFDLVVVEAFADYALVFGHLYDPA	
H	l....f.dilr...d.v.s.kkLm.k.Q...FDVVI.dAl..fg.llael...p	
M	l.....f.d.F.....D.v.s.keLmtk.Q...FDVll.dpiA..g.liael.q.p	
R	i.....f...y.....D.v.s.kqLmtk.Q...FDVl..dpiA..g.liael.h.p	
E	PVI---QIAPGYGLAENFDTVGAVER-HPVHHPNIW-RSNFDDTEANVTMRLYKEF	
H	.V...rfsPGYai..h.g.l...p...PV..sel..q..f.e...Nmi..v-LY.EF	
M	.l...rfsPGY.i...s.g...p...PV..s.l..q..f.e...Nmi..M-LY.df	
R	.l...fsPG..L...sig...p...PV..s.l..k..f.D...Nmi..M-LY.df	
E	--KIL-ANMSNALLKQFGPNTPTIEKLRNKVQLLLNLHPIDNNRPVPPSVQYLG	
H	.qIf..k..d.f.e..lG..T-Tl....K.di.Li....Fq..hPl.PnVefv-	
M	.qmF..k..dsf..e.lG..T-Tl....q.em.Li...n.le..hP..PnVdYv-	
R	.L...k..dtf..e.lG..T-Tvd...skVeI..Li...n.le..hP..PnVdYi-	
E	GGIHLVKSAPLTKLSPVINAQMKNKSGTIYVSFGSSIDTKSFANEFLYMLINTFKT	
H	GGIH...a.PL.K....f-.Q.s..ng..vf-SlGS.v--n.seE...vi.sal..	
M	GGIH...a.PL.K....f-.Q.s..g..vf-SlGS.v--n.teE...i.al..	
R	GGIH...a.PL.K....f-.Q.s..g..vf-SlGS.v--n.teE...i.al..	
Z	285 qp..G..YVSFGG....rp...Ei...L.ds...	
E	LDNYTILWKIDDEVVNITLPAVNITQWNFQRAVLRRHKMAAFITQGGLOSSDEAL	
H	i...-vLWrfDgn...l.L.t.l--kwI.Q..lL.H.K..AFhThGG.ng...Ai	
M	i...-vLWrfDg...l...t.v--kwI.Q..lL.H.K..AFhThGG.ng...EAI	
R	i...-vLWrfDg...l...t.v--kwI.Q..lL.H.K..AFvThGG.ng...EAI	
Z	L...-W.l...l...a.g..l..W..Q.AVLRH..vgAFvThaG..S..EgL	
E	EAGIPMVCLPMHMGQFYHAHKLQQLGVARALDVTVSSDQLLVAINDLVFNAPTYKK	
H	.p.IPMV.vPl.aDQ..n...mk..G.A.sLD..TmSS.dLL.Alk.Vi-N.P.YK.	
M	.GIPMi.iPl.GeQ..n...m..G.A.ALn..TmS..dvL.AleeVi--P.YKK	
R	.GIPMi.iPl.GDQ..n...m..G.A.sLn..TmS..dfL.AleeVi-d.P.YKK	
Z	.SGvPM.C.P..GDQ..nAr.v.hvG.G.Afe..ants..v..AveeL.....rr	
E	HMAELYALINHDKATFPPLDKAIKFTERVIRYRHDSRLQYSLKTTAANVPYSNYMM	
H	n...L-s.IhHDqp--PLDrA--F...v-.RH..akhL....A.dl....f.	
M	n...L-s.IhHDqp--PLDrA--F...v-.RH..akhL.pl---g.Nl....f.	
R	nv...L-s.IhHDqp--PLDrA--F...I-.RH..akhL.pl---g.NlP---Y.	
Z	..AEL.ALv.e..g.....K.f.f.F.E.V.R.*	
E	YKSVFSIVMNLTHF*	
H	Y.S-l.v....La.....	
M	Y.S-l.vi....Ls.....	
R	Y.S-l.vi....LT.F.....	



**Fig. 2.** AcMNPV-specific ecdysone UDP-glucosyl transferase activity. (A) *Spodoptera frugiperda* cells were infected with wt AcMNPV (wt) or vEGTZ at a multiplicity of infection of 20. Twelve hours later, the cells (C) and overlying medium (M) were harvested separately; the cells were lysed in tissue culture fluid by several strokes of a Dounce homogenizer. Mock-infected cell cultures were treated in parallel. The enzymatic assay was modified from Bansal and Gessner (16). The standard incubation mixture included the following: cell lysate containing 10  $\mu$ g of total protein or medium from the equivalent number of cells; 10 mM MgCl<sub>2</sub>; 10 mM tris maleate, pH 7.4; 1 mM unlabeled UDP-glucose (Sigma Chemical); 100  $\mu$ M unlabeled ecdysone (Sigma Chemical); and 0.25  $\mu$ Ci [<sup>3</sup>H]ecdysone (DuPont Biotechnology Systems). The final reaction volume was 50  $\mu$ l. Reactions were carried out for 5 min at 37°C, then stopped by the addition of two volumes of ethanol. Products were evaporated and resuspended in 60% ethanol. (B) Cell lysates from wt-infected cells were assayed as described above except that the concentration of UDP-glucose was reduced to 3.16  $\mu$ M and the reaction was allowed to proceed for 30 min. The UDP-glucose used was either unlabeled (lane 1) or UDP-[U-<sup>14</sup>C]glucose (0.05  $\mu$ Ci; DuPont Biotechnology Systems) (lane 2). [<sup>3</sup>H]Ecdysone ([<sup>3</sup>H]E) and UDP-[U-<sup>14</sup>C]glucose ([<sup>14</sup>C]G) were also subjected to chromatography. Autoradiographs of the silica-gel plates are presented. Scintillation counting confirmed the presence of both <sup>3</sup>H and <sup>14</sup>C in the ecdysone-glucose conjugate in (B), lane 2. The positions of unconjugated (E) and conjugated (EG) ecdysone, as well as UDP-glucose (UDPG), are indicated.

ecdysteroid inactivation mechanism (9), we tested whether AcMNPV infection, through the expression of *egt*, would disrupt the normal developmental process of the host insect. Sixteen newly ecdysed fourth instar *S. frugiperda* larvae were infected by injection either with wt AcMNPV or vEGTZ and monitored daily for any perturbations in their development. One cohort of 16 larvae was injected with tissue culture fluid as a negative control. All larvae injected with tissue culture fluid molted to fifth instar as expected. Only 1 of 16 larvae infected with wt virus made this transition. In contrast, all larvae infected with the mutant vEGTZ underwent a fourth-to-fifth instar molt. Thus, it is clear that *egt* expression by wt AcMNPV specifically inhibits host molting.

Both infected groups of larvae subsequently succumbed to the viral infection, showing that disruption of *egt* did not prevent vEGTZ from completing its infectious cycle. Using newly ecdysed fifth instar larvae, we have observed that no wt-infected larvae showed any signs of pupation, whereas the majority of vEGTZ-infected insects displayed several behavioral modifications (feeding cessation, wandering, and spinning) characteristic of an impending larval-pupal molt. However, all virus-infected animals died before pupation.

The data show that the baculovirus AcMNPV specifically interrupts the normal development of its insect host by producing an ecdysteroid UDP-glucosyl transferase, which is possibly secreted into the hemolymph by infected cells. Our observations provide the first genetic and developmental evidence in support of the hypothesis (9) that glucose conjugation by UDP-glucosyl transferases is a mechanism for ecdysteroid inactivation. AcMNPV *egt* is the first identified gene encoding a glucosyl transferase capable of conjugating ecdysteroids. The identification of an insect virus gene encoding such an enzyme should greatly facilitate the study of equivalent insect genes.

It is probable that other baculoviruses and other insect viruses also interfere with the development of their hosts by altering hormonal regulation. A more detailed analysis of the regulation of AcMNPV *egt* expression in the insect, as well as a closer study of the hormonal changes brought about by viral infection, will lead to a deeper understanding of the significance of this activity to the viral life cycle. In particular, such studies should reveal the evolutionary advantage gained by viruses expressing an ecdysteroid UDP-glucosyl transferase.

#### REFERENCES AND NOTES

1. Reviewed in R. G. Granados and B. A. Federici, Eds., *The Biology of Baculoviruses* (CRC Press, Boca

- Raton, FL, 1986), vols. I and II.
2. S. Kumar and L. K. Miller, *Virus Res.* 7, 335 (1987).
3. Reviewed in G. J. Dutton, in *Glucuronidation of Drugs and Other Compounds*, G. J. Dutton, Ed. (CRC Press, Boca Raton, FL, 1980), pp. 1-78.
4. Reviewed in J. N. Smith, in *Drug Metabolism: From Microbe to Man*, D. V. Parke and R. L. Smith, Eds. (Taylor and Francis, London, 1977), pp. 219-232.
5. E. J. Ralston, J. J. English, H. K. Doone, *Genetics* 119, 185 (1988).
6. M. R. Jackson et al., *Biochem. J.* 242, 581 (1987).
7. T. Kimura and I. S. Owens, *Eur. J. Biochem.* 168, 515 (1987).
8. P. I. Mackenzie, *J. Biol. Chem.* 262, 9744 (1987).
9. J. T. Warren, B. Steiner, A. Dorn, M. Pak, L. I. Gilbert, *J. Liq. Chromatogr.* 9, 1759 (1986); M. J. Thompson et al., *Arch. Insect Biochem. Physiol.* 4, 1 (1987); M. J. Thompson, J. A. Svoboda, R. Lozano, K. R. Wilzer, *ibid.* 7, 157 (1988).
10. H. H. Lee and L. K. Miller, *J. Virol.* 27, 754 (1978).
11. D. R. O'Reilly and L. K. Miller, unpublished data.
12. Reviewed in L. M. Riddiford, in *Comprehensive Insect Physiology, Biochemistry, and Pharmacology*, G. A. Kerkut and L. I. Gilbert, Eds. (Pergamon, Oxford, 1984), vol. 8, pp. 37-84; S. L. Smith, *ibid.*, vol. 7, pp. 295-341; J. Koolman and P. Karlson, *ibid.*, pp. 343-361.
13. The sequence has been deposited with GenBank, accession number M22619.
14. D. J. Lipman and W. R. Pearson, *Science* 227, 1435 (1985).
15. M. Dayhoff, *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Silver Spring, MD, 1978), vol. 5, suppl. 3.
16. S. K. Bansal and T. Gessner, *Anal. Biochem.* 109, 321 (1980).
17. We thank M. Brown for helpful discussions and advice on insect bioassays, as well as critical reading of the manuscript; A. Lea, R. Cummings, K. Spindler, and M. Adang for comments on the manuscript; I. Goldman and S. Hilliard for expert technical assistance; D. Perkins (USDA, Tifton, GA) for *S. frugiperda* eggs; and S. Karl and L. Passarelli for help constructing vEGTZ. This work was supported in part by Public Health Service grant AI 23719 from the National Institute of Allergy and Infectious Disease and by the University of Georgia Agriculture Experiment Station.

15 February 1989; accepted 6 July 1989

## T Cells Against a Bacterial Heat Shock Protein Recognize Stressed Macrophages

TETSUYA KOGA, ANGELA WAND-WÜRTTENBERGER, JACQUELINE DEBRUYN, MARTIN E. MUNK, BERND SCHOEL, STEFAN H. E. KAUFMANN\*

Heat shock proteins are evolutionarily highly conserved polypeptides that are produced under a variety of stress conditions to preserve cellular functions. A major antigen of tubercle bacilli of 65 kilodaltons is a heat shock protein that has significant sequence similarity and cross-reactivity with antigens of various other microbes. Monoclonal antibodies against this common bacterial heat shock protein were used to identify a molecule of similar size in murine macrophages. Macrophages subjected to various stress stimuli including interferon- $\gamma$  activation and viral infection were recognized by class I-restricted CD8 T cells raised against the bacterial heat shock protein. These data suggest that heat shock proteins are processed in stressed host cells and that epitopes shared by heat shock proteins of bacterial and host origin are presented in the context of class I molecules.

EXPOSURE OF CELLS TO VARIOUS stress conditions leads to the synthesis of a family of polypeptides termed heat shock proteins (hsp's) (1). After immunization with tubercle or leprosy bacilli, a major fraction of T cells and antibodies is directed against the 65-kD hsp, a newly discovered member of the hsp family (2-4). This hsp is a homolog of the GroEL hsp of 60 kD in *Escherichia coli* and shares some degree of homology with the *E. coli* DnaK hsp of 70 kD, which is a homolog of the mycobacterial 71-kD hsp (2). Pathogenic mycobacteria preferentially live and replicate

inside macrophages, and it can be assumed that the stress imposed by activated macrophages induces abundant hsp synthesis in their intracellular parasites, which will then result in a strong immune response to this antigen. T cells with specificity for the 65-kD hsp are not only demonstrable in patients but also in many healthy individuals, and a great variety of microbes are known to have similar hsp's (2, 3). Indeed, hsp's have been highly conserved in evolution, and in some cases similar molecules are shared by prokaryotic and eukaryotic cells (1). Our study reveals that a homolog of the mycobacterial 65-kD hsp is present in murine macrophages. This molecule is recognized by T cells raised against the mycobacterial 65-kD hsp and could serve as a target for protective and autoreactive immune responses.

Lysates of bone marrow-derived macro-

T. Koga, A. Wand-Württenberger, M. E. Munk, B. Schoel, S. H. E. Kaufmann, Department Medical Microbiology and Immunology, University of Ulm, D-7900 Ulm, Federal Republic of Germany.  
J. DeBruyn, Institut Pasteur du Brabant, B-1180 Bruxelles, Belgium.

\*To whom correspondence should be addressed.

# INHERITANCE OF CYCLIC HYDROXAMATES IN *Zea mays* L.

G. M. DUNN, B. J. LONG, and D. G. ROUTLEY

Department of Plant Science, University of New Hampshire, Durham, N.H. 03824.  
New Hampshire Agricultural Experiment Station Contribution no. 1011, received 17  
Dec. 1980, accepted 15 Mar. 1981.

DUNN, G. M., LONG, B. J. AND ROUTLEY, D. G. 1981. Inheritance of cyclic  
hydroxamates in *Zea mays* L. Can. J. Plant Sci. 61: 583-593.

Hydroxamic acids have been implicated in the resistance of corn (*Zea mays* L.) to both fungi and insects. In this study, five selected crosses were used among the four inbreds *BxBx*, *bxbx*, B49 and B37 to study inheritance of hydroxamates. Hydroxamate concentration in the parental,  $F_1$ ,  $F_2$  and backcross generations for each cross was estimated by a rapid procedure based upon the colorimetric reaction of hydroxamates with  $FeCl_3$ . Components of variance and estimates of heritability were obtained by the procedures of Warner (1952).  $F_2$  and backcross data indicated that concentration of hydroxamates is controlled monogenically in the cross *bxbx*  $\times$  *BxBx* and polygenically in the crosses *bxbx*  $\times$  B49 and *bxbx*  $\times$  B37. Estimates of gene number using the Castle-Wright formula indicated that hydroxamate concentration is conditioned in B49 and B37 by five and two loci, respectively. The addition of *BxBx* to either B49 or B37 increased the frequency of genotypes in  $F_2$  with a high concentration of hydroxamates. Additive genetic variance was the most important component of the phenotypic variance and resulted in estimates of heritability from 0.64 to 0.79. However, the dominance component of variance was considerably higher for crosses involving *BxBx* than for the crosses *bxbx*  $\times$  B49 and *bxbx*  $\times$  B37.

Les acides hydroxamiques semblent jouer un rôle dans la résistance du maïs (*Zea mays* L.) aux champignons cryptogames et aux insectes. Dans l'étude décrite ci-après, nous avons examiné le mode de transmission héréditaire de ces substances sur cinq croisements réalisés entre quatre lignées consanguines: *BxBx*, *bxbx*, B49 et B37. La concentration en hydroxamates chez les lignées parentales, les  $F_1$  et  $F_2$  et les rétrocroisements impliqués dans chaque croisement a été estimée par une méthode rapide utilisant la réaction colorimétrique des hydroxamates au  $FeCl_3$ . L'analyse de la variance et les estimées de l'héritabilité ont été calculées par les méthodes de Warner (1952). Les données obtenues sur les  $F_2$  et les rétrocroisements montrent que la concentration en hydroxamates est modulée par un seul gène dans le croisement *bxbx*  $\times$  *BxBx* et par des gènes multiples dans les croisements *bxbx*  $\times$  B49 et *bxbx*  $\times$  B37. Les estimées du nombre de gènes, établies par la formule de Castle-Wright, révèlent que ce caractère est gouverné par cinq loci dans B49 et par deux dans B37. L'adjonction de *BxBx* à B49 ou B37 a donné lieu à une fréquence accrue de génotypes  $F_2$  à forte teneur en hydroxamates. La variance génétique additive était la composante la plus importante de la variance phénotypique et a produit des estimées de l'héritabilité allant de 0.64 à 0.79. Toutefois, l'élément de dominance dans la variance était considérablement plus élevé pour les croisements comportant *BxBx* que pour les croisements *bxbx* avec B49 ou B37.

The cyclic hydroxamate 2,4-dihydroxy-7-methoxy-1, 4 benzoxazin-3-one (DIMBOA) was first reported in corn (*Zea mays* L.) and wheat (*Triticum aestivum* L.) (Wahlroos and Virtanen 1959). Subsequent studies have im-

plicated the hydroxamates in resistance to stalk rot (BeMiller and Pappellis 1965), northern corn leaf blight (Anglade and Molot 1967; Long et al. 1975), European corn borer (Klun et al. 1967), and corn leaf aphid (Long et al. 1977).

Can. J. Plant Sci. 61: 583-593 (July 1981)



Monogenic and polygenic control have been hypothesized for accumulation of hydroxamates in corn. Hamilton (1964) studied the inheritance of hydroxamates in corn seedlings. The presence of DIMBOA or its glucoside was indicated by a blue-colored chelate formed when a root tip was crushed on filter paper impregnated with 0.1 M  $\text{FeCl}_3$ . He obtained a 3:1 ratio in  $F_2$  for high versus low hydroxamate seedlings. This gene conditioning production of DIMBOA was designated as *BxBx* (Couture et al. 1970). Klun et al. (1970) presented evidence for quantitative inheritance of DIMBOA. General combining ability accounted for 91% of the variation among hybrids in DIMBOA concentration.

The objectives of this study were to estimate gene number controlling concentration of hydroxamates in selected corn inbreds, to determine its heritability, and to estimate components of variance in five crosses.

### MATERIALS AND METHODS

The following inbreds were used in this study: B49 — high DIMBOA line, polygenic; B37 — low DIMBOA line, polygenic; *BxBx* — high DIMBOA line, monogenic; and *bxbx* — low DIMBOA line, monogenic. The first two inbreds were obtained from Dr. A. R. Hallauer, Iowa State University. The latter two lines were obtained originally, as genetic stocks from Dr. R. H. Hamilton, Pennsylvania State University and have since been selfed for six to seven generations each.

Five separate crosses were made: *bxbx*  $\times$  *BxBx*; *bxbx*  $\times$  B49; *bxbx*  $\times$  B37; B49  $\times$  *BxBx*; and B37  $\times$  *BxBx*. Each parent ( $P_1$ ,  $P_2$ ) and the corresponding  $F_1$ ,  $F_2$ ,  $BC_1$  ( $F_1 \times P_1$ ) and  $BC_2$  ( $F_1 \times P_2$ ) were analyzed for concentration of hydroxamates using the rapid procedure (Long et al. 1974).

Approximately 50 plants of each parent and  $F_1$ , 100 plants of each backcross and 250 plants in each  $F_2$  were analyzed for hydroxates. All plants were grown in the greenhouse at approximately 15°C (night) and 25°C (day) on a 16-h photoperiod to eliminate gross environmental effects.

The rapid procedure consisted of removing 0.1- to 0.2-g sections of tissue from the lower stems of seedlings 36-41 cm in height (5th to 6th leaf stage). The samples were placed in plastic vials and frozen

overnight. Subsequent thawing of the tissue allowed the B-glucosidase to hydrolyze the glucoside and release DIMBOA. Crude extracts were prepared by crushing the tissue with mortar and pestle in 1.0 mL of a solution containing equal volumes of 95% ethanol and 0.1 N HCl. The extract was centrifuged and decanted into a cuvette to which was added 0.1 mL of 0.1 N  $\text{FeCl}_3$ . Absorbance values were obtained and concentrations determined from a standard curve prepared from purified DIMBOA.

The rapid procedure based upon the reaction to  $\text{FeCl}_3$  is not specific for DIMBOA. Although DIMBOA is by far the major hydroxamate in *Zea mays*, it is likely that other related hydroxamates are present in trace quantities. For this reason, the data are expressed as total hydroxamate concentration rather than as concentration of DIMBOA.

Estimates of genetic effects and components of genetic variance were obtained using the procedure outlined by Warner (1952). The notation has been modified as indicated in the text. In this study, the environmental variance was obtained from the average of  $P_1$ ,  $P_2$  and  $F_1$  variances since each was considered genetically uniform. The additive component of variance  $12 V_A$ , was estimated by multiplying  $V_{F_1}$  by two and subtracting from it the combined backcross variances, thus eliminating the dominance  $14 V_D$  and environmental components  $V_E$ . By subtracting the additive and environmental variances from the total phenotypic variance of the  $F_2$ , an estimate of the dominance variance was obtained. The degree of dominance was calculated as  $\sqrt{4V_D/2V_A}$ . Estimates of heritability (narrow sense) were calculated according to the equation  $h^2 = \frac{1}{2}V_A(\frac{1}{2}V_A + \frac{1}{4}V_D + V_E)$ . Estimates of gene number were calculated according to Castel's formula  $N = \frac{D^2}{8(V_{F_2} - V_{F_1})}$  where  $N$  = gene number,  $D$  = the difference between the parental means, and  $V_{F_1}$  and  $V_{F_2}$  = the phenotypic variances of the  $F_1$  and  $F_2$ , respectively. All values were transformed to logarithms in order to satisfy the assumption of scaling upon which the analysis is based. The frequency distributions are based on the log of the hydroxamate concentration whereas the range and mean values in Table 1 are given in terms of mg hydroxamates/g fresh weight. Additional details are given elsewhere (Long 1977).

### RESULTS AND DISCUSSION

Ranges, means and variances for the parents and for five crosses among them are given in



Table 1. The parents differed widely in concentrations of hydroxamates, with *bxbx* (monogenic) and B49 (polygenic) representing the extremes.

In the cross *bxbx* × *BxBx*, the average concentrations of hydroxamates in  $F_2$  and  $F_3$  exceeded the mid-parental value of 0.538 mg, indicating the effects of dominance. The backcrosses produced concentrations nearly intermediate to the parental values. Variances for concentration of hydroxamates were considerably larger for segregating than for non-segregating generations.

The frequency distributions (Fig. 1) indicated incomplete dominance for this trait in

this cross. Distribution in the  $F_2$  appeared to be trimodal. When concentration of hydroxamates was grouped into discrete classes, a good fit was obtained to a 1:2:1 ratio in  $F_2$  and to a 1:1 ratio in the backcrosses.  $F_3$  data tended to support the hypothesis that a single, partially dominant gene conditions much of the hydroxamates in *BxBx*.

The  $F_1$  and  $F_2$  means were fairly close to the mid-parental value of 0.595 mg in the cross *bxbx* × B49. Mean concentrations of hydroxamate in each backcross closely approximated the mid-parental values of 0.369 and 0.884 mg for  $BC_1$  and  $BC_2$ , respectively. Frequency distributions for  $F_2$ ,  $BC_1$

Table 1. Ranges, means and variances for hydroxamate concentration in parents and progenies from five crosses in *Zea mays*

Population	No. of plants	Range	Mean (mg/g fresh wt)	Variance†
Parents				
<i>bxbx</i>	51	0.022-0.133	0.081	0.000149
<i>BxBx</i>	48	0.753-1.172	0.996	0.000606
B37	50	0.254-0.441	0.333	0.000241
B49	49	0.843-1.271	1.110	0.000406
Crosses				
<i>bxbx</i> × <i>BxBx</i>				
$F_1$	50	0.533-0.901	0.663	0.000448
$F_2$	241	0.002-1.142	0.556	0.008601
$F_3$	133	0.077-1.110	0.552	0.008320
$BC_1$ ‡	113	0.011-0.878	0.389	0.008292
$BC_2$ ‡	107	0.578-1.198	0.812	0.002080
<i>bxbx</i> × B49				
$F_1$	50	0.454-0.811	0.658	0.000360
$F_2$	245	0.164-1.010	0.545	0.002314
$BC_1$	96	0.122-0.677	0.382	0.001850
$BC_2$	104	0.619-1.100	0.864	0.000830
<i>bxbx</i> × B37				
$F_1$	50	0.164-0.293	0.232	0.000230
$F_2$	239	0.064-0.444	0.201	0.000682
$BC_1$	97	0.013-0.384	0.149	0.000431
$BC_2$	107	0.123-0.446	0.288	0.000464
B49 × <i>BxBx</i>				
$F_1$	50	0.962-1.263	1.069	0.000365
$F_2$	242	0.296-1.373	0.884	0.003660
$BC_1$	102	0.351-1.424	0.913	0.002976
$BC_2$	104	0.636-1.409	0.931	0.001910
B37 × <i>BxBx</i>				
$F_1$	50	0.562-0.844	0.729	0.000234
$F_2$	241	0.154-1.053	0.552	0.004212
$BC_1$	103	0.263-1.111	0.645	0.003460
$BC_2$	105	0.500-1.206	0.816	0.001856

†Original data were transformed to logarithms for calculation of variances.

‡In each cross,  $BC_1$  is backcross of  $F_1$  to female parent (first parent listed), and  $BC_2$  is backcross to male parent.

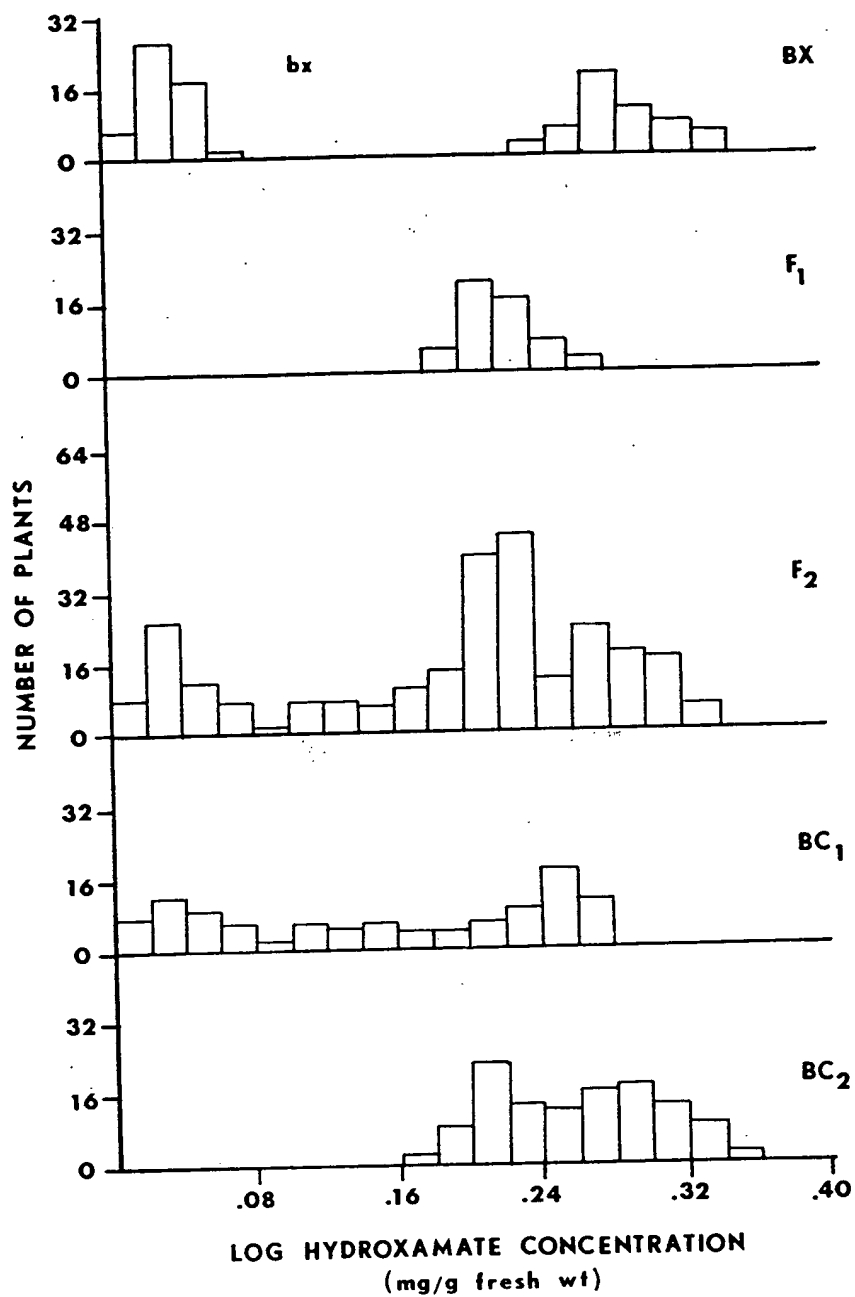


Fig. 1. Frequency distributions for the parental,  $F_1$ ,  $F_2$ ,  $BC_1$  and  $BC_2$  populations for hydroxamate concentration from the cross  $bx \times BxBx$ .

and  $BC_2$  were continuous and seemed to follow a normal distribution curve (Fig. 2). It was not possible to divide concentration of hydroxamates in  $F_2$  of  $bxbx \times B49$  into discrete classes as was the case for  $bxbx \times BxBx$ . From these data it appears that concentration of hydroxamates is controlled in  $bxbx \times B49$  by several genes having a small but cumulative effect. It is hypothesized that B49 does not possess any one gene with an effect comparable to  $BxBx$ . Estimates of gene number using the Castle-Wright formula indicated that five loci condition concentration of hydroxamates in B49.

Both parents had low concentration of hydroxamates in the  $bxbx \times B37$  cross. Concentrations of hydroxamates in  $F_1$  and  $F_2$  were close to the mid-parental value of 0.207 mg. Backcross means were also close to the mid-parental values. Variances in  $F_2$ ,  $BC_1$ , and  $BC_2$  were considerably larger than for parental and  $F_1$  generations, but smaller than those for  $bxbx \times B49$ . These lower variances in the  $bxbx \times B37$  cross probably reflected the level of hydroxamates in the parents.

As in the  $bxbx \times B49$  cross, the frequency distributions for the  $F_2$ ,  $BC_1$  and  $BC_2$  populations were continuous and normally distributed (Fig. 3). The data indicated that content of hydroxamates in B37 is controlled by genes with little or no dominance. The estimate of gene number indicated that two loci condition concentration of hydroxamates in B37.

In the cross  $B49 \times BxBx$ , both parents exhibited high mean concentrations of hydroxamates. Mean concentration in the  $F_1$  was nearly intermediate to that of the parents. The frequency distribution in  $F_2$  was highly skewed to the right (Fig. 4). This departure from normality may reflect the influence of  $BxBx$  on the polygenic expression of hydroxamate concentration in B49.

Backcrosses of the  $F_1$  to either parent resulted in similar mean concentrations of hydroxamates. The distribution for  $BC_1$  ( $F_1 \times B49$ ) clearly exhibited two modes, while in  $BC_2$  ( $F_1 \times BxBx$ ) the two modes were less discernible. The bimodal nature of the back-

cross distributions can be explained in part by considering the  $Bx$  gene alone. Since B49 probably does not possess the  $Bx$  gene, backcrosses of the  $F_1$  to the two parents would result in the expected genotypes  $BxBx:Bxbx$  for the  $F_1$  backcrossed to  $BxBx$  and  $BxBx:bxbx$  for the  $F_1$  backcrossed to B49. The addition of polygenes from B49 plus environmental effects would tend to blur these distinct differences.

In the  $B37 \times BxBx$  cross, the relatively large deviation of the  $F_1$  mean from the mid-parental value of 0.664 mg indicates the effects of dominance. As in the previous cross, the  $F_2$  frequency distribution appeared highly skewed to the right (Fig. 5). The effect of the  $Bx$  gene was shown by the departure of the  $F_2$  distribution from normality. Similar trends were apparent in the two backcross populations. The frequency distribution for both  $BC_1$  and  $BC_2$  were bimodal.

Components of variance for concentrations of hydroxamates in the five crosses are shown in Table 2. The phenotypic variances were considerably higher for crosses involving  $BxBx$  than for those with  $bxbx$ . This was probably a reflection of the dominance effects of  $BxBx$  which resulted in a higher frequency of extreme genotypes in the  $F_2$  populations. Similarly, the smaller variances shown by the crosses  $bxbx \times B49$  and  $bxbx \times B37$  reflect the normal distribution curves in which the majority of individual plants were clustered about the means.

The environmental variances for each cross were small in comparison to the total phenotypic variances. Genetic variance therefore accounted for the majority of the phenotypic variation for this trait. In all crosses the additive genetic variance was considerably higher than the dominance variance and accounted for the majority of the genetic variation. However, dominance variation was proportionally higher in the crosses involving  $BxBx$ .

Estimates of heritability, degree of dominance and percent dominance for concentration of hydroxamates in the five crosses are shown in Table 3. Estimates of heritability

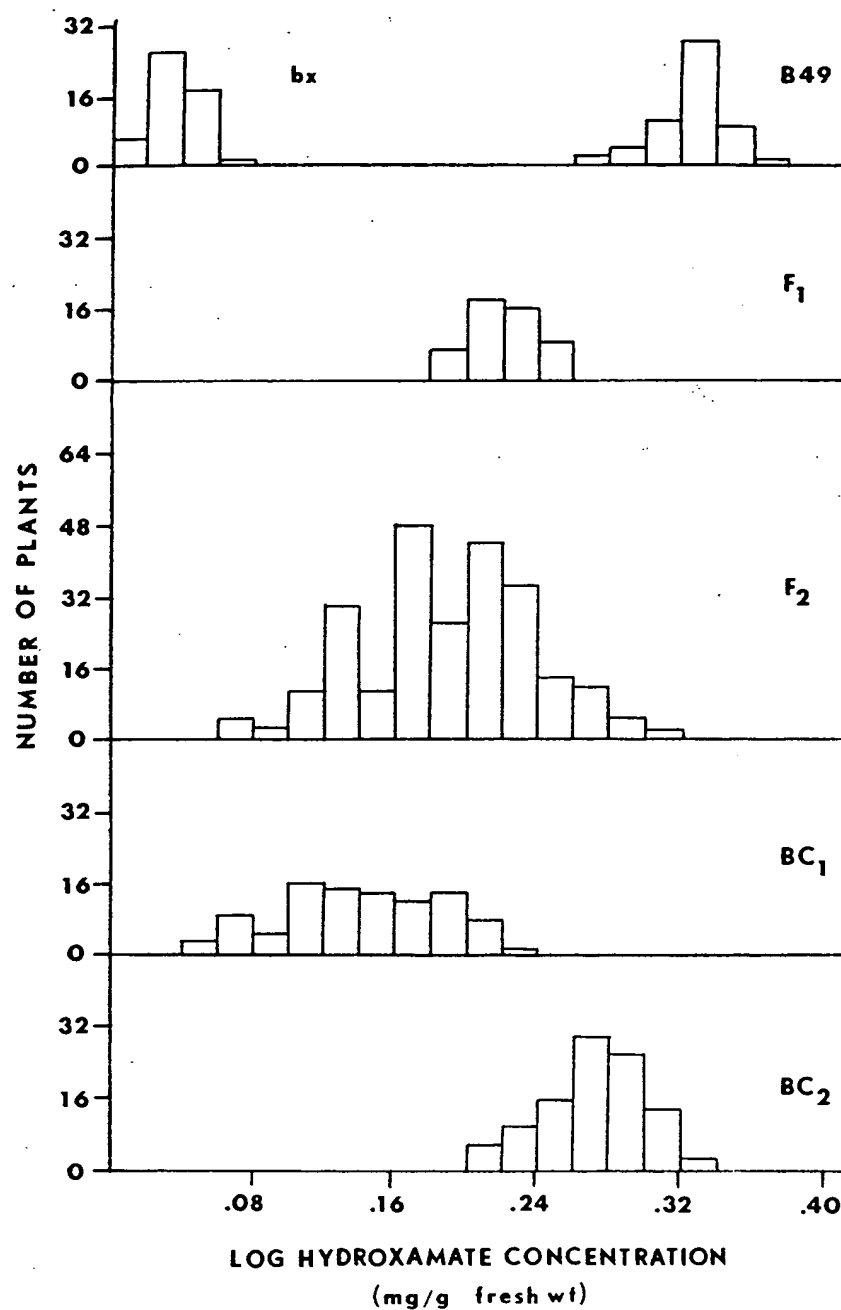


Fig. 2. Frequency distributions for the parental, F<sub>1</sub>, F<sub>2</sub>, BC<sub>1</sub> and BC<sub>2</sub> populations for hydroxamate concentration from the cross *bx* × B49.

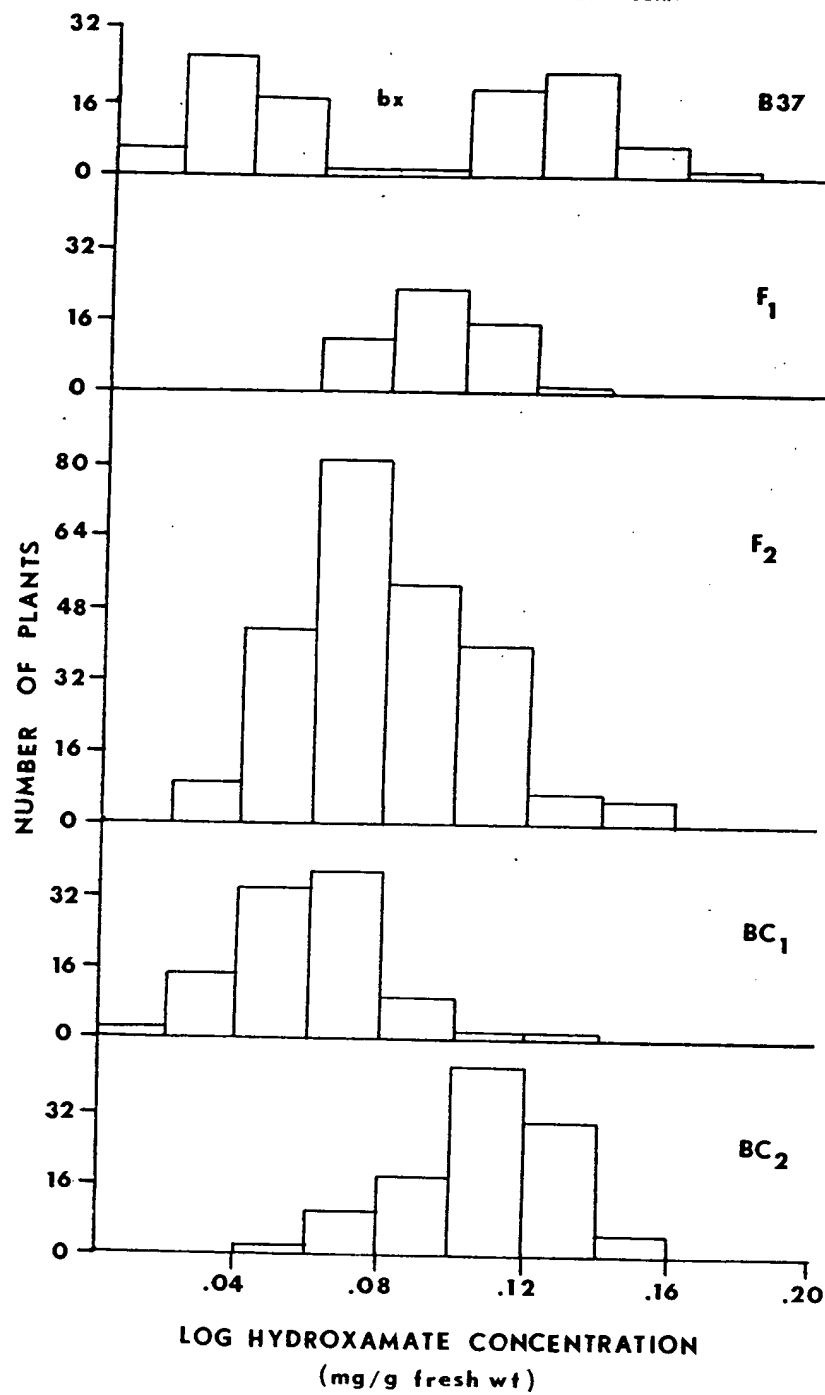


Fig. 3. Frequency distributions for the parental, F<sub>1</sub>, F<sub>2</sub>, BC<sub>1</sub> and BC<sub>2</sub> populations for hydroxamate concentration from the cross *bxb* × B37.

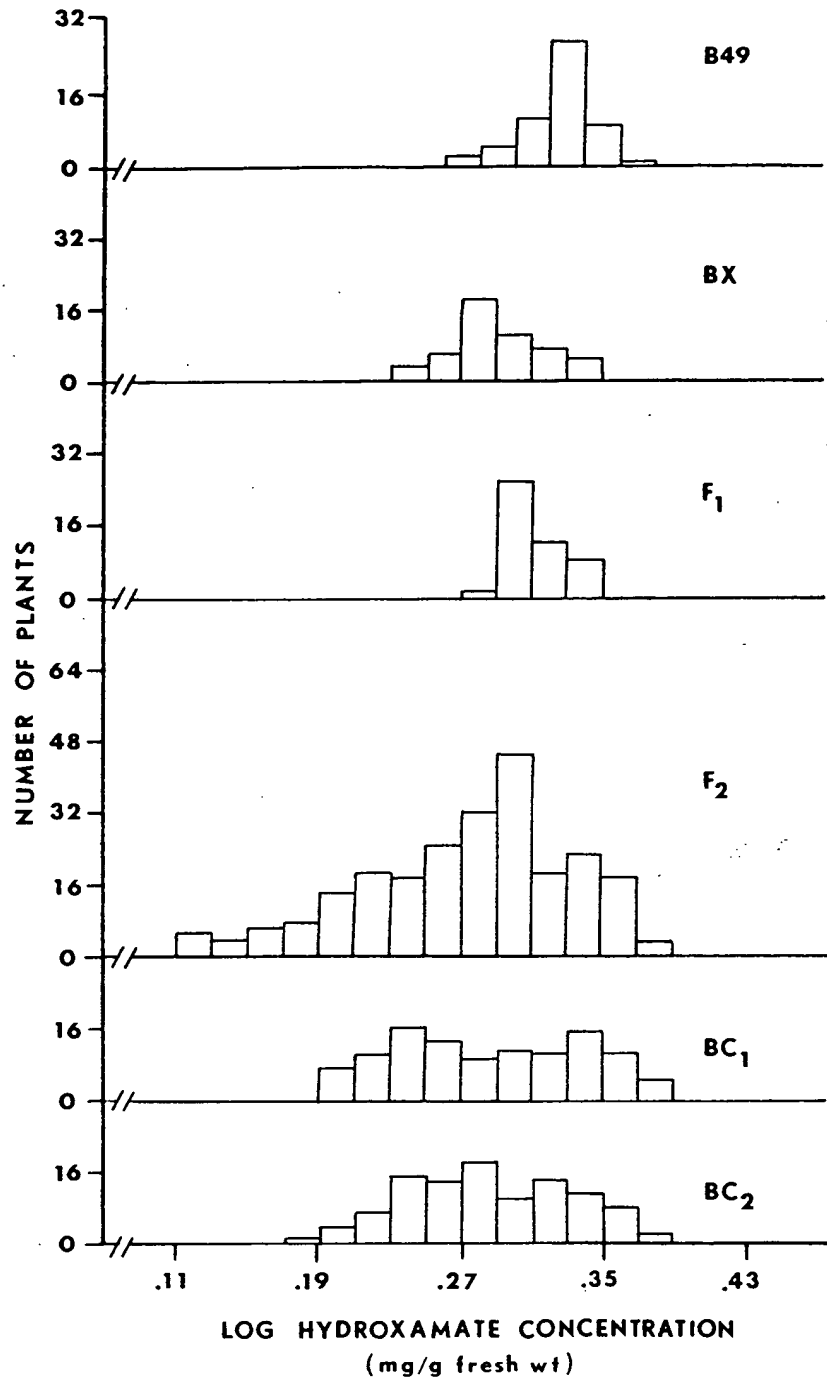


Fig. 4. Frequency distributions for the parental, F<sub>1</sub>, F<sub>2</sub>, BC<sub>1</sub> and BC<sub>2</sub> populations for the hydroxamate concentration from the cross B49  $\times$  BxBx.

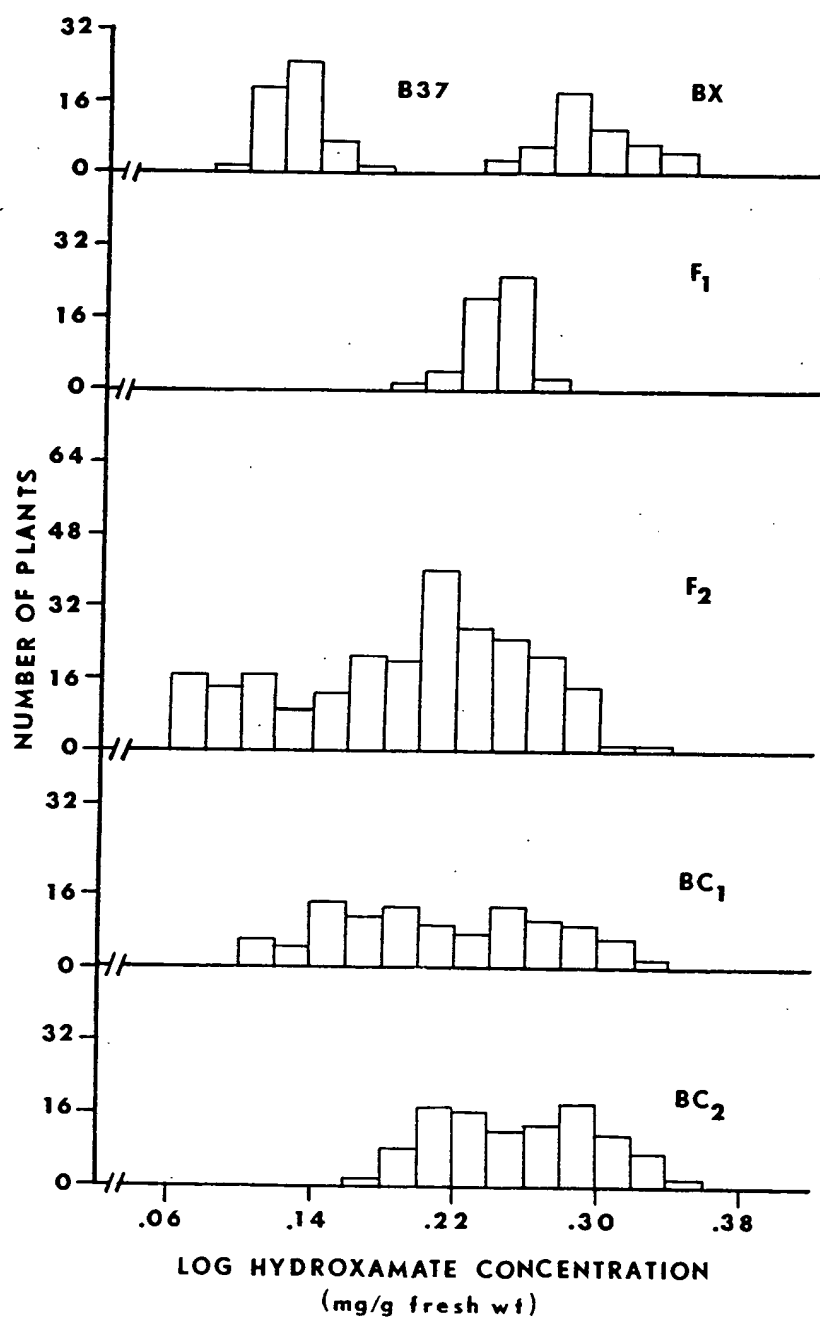


Fig. 5. Frequency distributions for the parental, F<sub>1</sub>, F<sub>2</sub>, BC<sub>1</sub> and BC<sub>2</sub> populations for hydroxamate concentration from the cross B37 × BxBx.

Table 2. Components of variance for hydroxamate concentration in selected crosses among four maize inbreds

Cross†	V <sub>P</sub>	V <sub>A</sub>	V <sub>D</sub>	V <sub>E</sub>
<i>bxbx</i> × <i>BxBx</i>	8.601‡	6.830	1.370	0.401
<i>bxbx</i> × B49	2.314	1.950	0.059	0.305
<i>bxbx</i> × B37	0.682	0.468	0.007	0.210
B49 × <i>BxBx</i>	3.660	2.434	0.767	0.459
B37 × <i>BxBx</i>	4.212	3.104	0.748	0.360

†Original data transformed to logarithms for analysis.

‡All values × 10<sup>-3</sup>.

ranged from 0.64 to 0.79 and reflect the high proportion of additive genetic variation to total phenotypic variation. Based upon these estimates, selection for concentration of hydroxamates by nearly any breeding method should be effective in these crosses.

It is apparent from Table 3 that crosses involving *BxBx* exhibited a much higher degree of dominance than in the crosses *bxbx* × B49 and *bxbx* × B37. It is hypothesized that the inbreds B49 and B37 do not possess any gene or genes with dominance effects comparable to *BxBx*.

Similarly, percent dominance values were comparably higher for crosses involving *BxBx* than for crosses with *bxbx*. This is a reflection of the higher proportion of dominance variation to total phenotypic variation shown by the *BxBx* crosses.

These data support the hypothesis that concentration of hydroxamates in *BxBx* is conditioned primarily by a major, partially dominant gene, while the concentration in B37 and B49 appears to be conditioned by

Table 3. Estimates of heritability, degree of dominance, and percent dominance for hydroxamate concentration in selected crosses among four maize inbreds

Cross†	Heritability	Degree of dominance	Percent dominance to the total variation
<i>bxbx</i> × <i>BxBx</i>	0.79	0.63	16.71
<i>bxbx</i> × B49	0.64	0.25	2.95
<i>bxbx</i> × B37	0.68	0.17	1.47
B49 × <i>BxBx</i>	0.67	0.79	23.96
B37 × <i>BxBx</i>	0.74	0.69	19.42

†Original data transformed to logarithms for analysis.

genes with smaller effects. A recurrent selection program is in progress to increase concentration of hydroxamates in corn utilizing both monogenic and polygenic factors.

#### ACKNOWLEDGMENTS

Appreciation is extended to Pioneer Hi-Bred International, Inc. for support for this work.

ANGLADE, P. and MOLOT, P. M. 1967. Mise en évidence d'une relation entre les sensibilités de lignées de maïs à la pyrale (*Ostrinia nubilalis* (HBN.)) et à l'hilinthasporiose (*Helminthosporium turcicum* Pass.). Ann. Epiphyties 18: 279-284.

BEMILLER, J. N. and PAPPELIS, A. J. 1965. 2,4-Dihydroxy-7-methoxy-1,4-benzoxazin-3-one glucoside in corn. I. Relation of water soluble, 1-butanol-soluble glycoside fraction content of pith cores and stalk rot resistance. II. Isolation of 6-methoxy-2(3)-benzoxazolinone fraction as a measure of glucoside content and tissue difference of glucoside content. Phytopathology 55: 1237-1243.

COUTURE, R. M., ROUTLEY, D. G. and DUNN, G. M. 1970. Role of cyclic hydroxamic acids on monogenic resistance to *Helminthosporium turcicum* in maize. Maize Genet. Coop. News Lett. 44: 169-170.

HAMILTON, R. H. 1964. A corn mutant deficient in 2,4-dihydroxy-7-methoxy-1, 4-benzoxazin-3-one with an altered tolerance of atrazine. Weeds 12: 27-31.

KLUN, J. A., GUTHRIE, W. D., HALLAUER, A. R. and RUSSELL, W. A. 1970. Genetic nature of the concentration of 2,4-dihydroxy-7-methoxy 2H-1, 4-benzoxazin-3(4H)-one and resistance to the European corn borer in a diallel set of eleven maize inbreds. Crop Sci. 10: 87-90.

KLUN, J. A., TIPTON, C. L. and BRINDLEY, T. A. 1967. 2,4-dihydroxy-7-methoxy-1,4 benzoxazin-3-one (DIMBOA), an active agent in the resistance of maize to the European corn borer. J. Econ. Entomol. 60: 1529-1533.

LONG, BRYANT J. 1977. Breeding and genetics of hydroxamates in *Zea mays* L. Ph.D. Thesis, Univ. of New Hampshire, Durham, N.H.

LONG, B. J., DUNN, G. M., BOWMAN, J. S. and ROUTLEY, D. G. 1977. Relationship of hydroxamic acid content in maize and resistance to the corn leaf aphid. Crop Sci. 17: 55-88.

LONG, B. J., DUNN, G. M. and ROUTLEY, D. G. 1974. Rapid procedure for estimating cyclic



hydroxamate (DIMBOA) concentration in maize. *Crop Sci.* **14**: 601-603.

LONG, B. J., DUNN, G. M. and ROUTLEY, D. G. 1975. Relationship of hydroxamic acid content and resistance to northern corn leaf blight. *Crop Sci.* **15**: 333-335.

WAHLROOS, I. and VIRTANEN, A. I. 1959. The precursors of 6-methoxy-benzoxazolinone in maize and wheat plants, their isolation and some of their properties. *Acta Chem. Scand.* **13**: 1906-1908.

WARNER, J. N. 1952. A method for estimating heritability. *J. Agron.* **44**: 427-430.

Claim 64  
Pathogenesis related  
proteins

# PLANT PATHOGENESIS-RELATED PROTEINS INDUCED BY VIRUS INFECTION

*J. F. Bol and H. J. M. Linthorst*

Department of Biochemistry, Gorlaeus Laboratories, Leiden University, Einsteinweg  
5, 2333 CC Leiden, The Netherlands

*B. J. C. Cornelissen*

MOGEN International, Einsteinweg 97, 2333 CB Leiden, The Netherlands

KEY WORDS: pathogenesis-related proteins, hypersensitive response, induced resistance,  
beta-1,3-glucanases, chitinases

## INTRODUCTION

The hypersensitive response triggered by an incompatible plant-pathogen interaction is accompanied by numerous metabolic changes (for reviews see 5, 8, 20, 37, 119). One general effect is the induction of enzymes that are involved in the synthesis of ethylene, a plant hormone able to induce many stress responses. The induction of enzymes from the phenylpropanoid pathway, oxidative enzymes, and hydroxyproline-rich glycoproteins (HRGP; extensin) is partially related to cell wall modifications that lead to the formation of a physical barrier to further spread of the pathogen from the site of infection. Peroxidases are involved in the polymerization of alcohol derivatives of aromatic compounds such as coumaric, ferulic, and sinapic acids into lignin and suberin and the cross-linking of these polyphenols to extensin molecules from the cell wall matrix. In addition to inducing a defense aimed at localizing the pathogen, various antimicrobial compounds are synthesized. These compounds include phytoalexins synthesized along branches of the phenylpropanoid pathway leading to furano-coumarins or isoflavonoids.

hydrolytic enzymes that may attack polysaccharide components from cell walls of the pathogen, and proteinase inhibitors that specifically block the digestive enzymes of herbivores (see Ryan, this volume). Collectively, these induced reactions are thought to contribute to an active defense mechanism of the plant known as "induced resistance."

Similar metabolic changes are induced by the incompatible interaction of a plant with either a fungus, a bacterium, or a virus (20, 37, 119). In this chapter we focus on the "pathogenesis-related" (PR) proteins that are induced in plants after infection with necrotizing viruses. In the early 1970s, Van Loon & Van Kammen (121) and Gianinazzi et al (42) independently reported the de novo synthesis of several proteins in tobacco plants reacting hypersensitively to infection with tobacco mosaic virus (TMV). These proteins were characterized by their acidic nature (43, 115), their resistance to proteases (117), and their extracellular location (88). More recently, basic homologues to a number of acidic PR proteins have been identified. In tobacco, the basic PR proteins are not secreted into the intercellular space of the leaf but accumulate in the vacuoles. Table 1 lists over 20 acidic and basic PR proteins purified to homogeneity from TMV-infected Samsun NN tobacco by Fritig et al (38, 53, 55). A number of acidic PR proteins have also been purified by Antoniwi et al (2), Pierpoint (94) and Van Loon et al (120). Different nomenclatures are used by various research groups (6); here, we adhere to that used by Fritig et al (38).

The tobacco PR proteins listed in Table 1 have been classified into five groups (120). Group 1 consists of the PR-1 proteins of unknown function. A serological cross-reactivity between the acidic and basic PR-1 proteins has been reported (55). Group 2 contains PR proteins with  $\beta$ -1,3-glucanase activity, which was detected first for PRs 2, N and O (56) and later for Q' and O' (38). The TMV-induced basic tobacco  $\beta$ -1,3-glucanase (Gluc. b) is probably identical to a similar enzyme produced in tobacco tissue cultured on auxin-containing medium (33, 104). PRs 2, N, O, Q' and Gluc. b are serologically related but antisera to these proteins do not cross-react with PR-O' (38). PR-O', which occurs as a dimer, is less strongly induced after TMV infection than the other PR proteins. In an in vitro assay, the enzymatic activities of PR-O and Gluc. b have are similar whereas that of the other  $\beta$ -1,3-glucanases are 20- to 100-fold lower (38, 56).

TMV-infection was found to induce four tobacco chitinases, namely the acidic PR proteins P and Q and the basic proteins Ch. 32 and Ch. 34 (66). The basic enzymes had been identified earlier in cultured tobacco tissue (103). All four proteins are serologically related and constitute group 3 (Table 1). Although the basic chitinases are six times more active than their acidic isozymes in an in vitro assay, the more acidic forms (PRs P and Q) were estimated to account for one third of the TMV-induced chitinase activity in

Table 1 PR proteins induced in Samsun tobacco (NN genotype) by TMV infection

Group	Name <sup>a</sup>	Acidic PR proteins		Basic PR proteins		Function
		Name	Mol wt (kd)	Name	Mol wt (kd)	
1		1a	15.8	16 kd	16.0	Unknown
		1b	15.5			
		1c	15.6			
2a		2	39.7	Gluc.b	33.0	$\beta$ -1, 3-Glucanase
		N	40.0			
		O	40.6			
		Q'	36.0			
2b		O'	25.0			$\beta$ -1, 3-Glucanase
3		P	27.5	Ch.32	32.0	Chitinase
		Q	28.5	Ch.34	34.0	
4		s1	14.5			Unknown
		r1	14.5			
		s2	13.0			
		r2	13.0			
5a		R	24.0	Osmotin	24.0	Unknown (TL-proteins <sup>b</sup> )
		S	24.0			
5b				45 kd	45.0	Unknown

<sup>a</sup> Nomenclature is according to Fritig et al (38)

<sup>b</sup> Thaumatin-like proteins

tobacco (66). Chitinases and  $\beta$ -1,3-glucanases are thought to play a role in a plant's defense mechanism against fungal infection (102).

The low molecular weight proteins classified as group 4 have been characterized in less detail. Two TMV-induced proteins with molecular weights of approximately 13 kd and 15 kd were recognized by Pierpoint (94) and Van Loon et al (120). Fritig et al (38) and Kaufmann (55) purified two 13.0-kd proteins (s2 and r2) and two 14.5-kd proteins (s1 and r1); r1 and r2 comigrate, as do s1 and s2, in alkaline nondenaturing polyacrylamide gels. The serological relationship between proteins of groups 2a and 4 reported by Van Loon et al (120) was not found by Kaufmann (55).

Two acidic 24-kd proteins are induced after TMV infection of *Nicotiana tabacum* L., cv Samsun NN (55) or *Nicotiana tabacum* L. cv Xanthi nc (95) tobacco; these are named PR-R and PR-S. In addition, a TMV-induced basic 24-kd protein has been identified that is serologically related to PRs R and S

(55). This latter protein is probably identical to osmotin, a protein induced by salt stress of tobacco (105). All three proteins, placed in group 5a (Table 1), show a 65% amino acid sequence homology to the sweet-tasting protein thaumatin (23, 95, 105). A similar degree of homology was found with a maize protein that is a bifunctional inhibitor of  $\alpha$ -amylase and proteases of insects (97), suggesting that the tobacco proteins play a role in plant defense against insects. However, purified PR-R and PR-S did not inhibit  $\alpha$ -amylase or various proteases in an *in vitro* assay (B. Fritig, personal communication).

The purification of a TMV-induced 45-kd tobacco protein has recently been reported (55) and, because of its serological cross-reactivity to PR-R and PR-S, has tentatively been placed in group 5b. Furthermore, a serological relationship between PR-R and PR-S and PR proteins of group 1 has been reported (120), but this relationship is not paralleled by a detectable amino acid sequence similarity.

Proteins related by serology or amino acid sequence similarity to the TMV-induced tobacco proteins from groups 1, 2a, 3, and 5a occur in over 20 plant species, including monocots and dicots (5, 6, 118). This indicates that these proteins, which may constitute up to 10% of the total soluble protein under various stress conditions, are highly conserved in the plant kingdom. Below, the current state of knowledge about these proteins and their corresponding genes is reviewed.

## INDUCTION OF PR PROTEINS

Similar patterns of proteins are induced after the incompatible interaction of plants with different types of pathogens and each interaction, in turn, induces a broad range of defense reactions. For example, after infection with TMV, tobacco plants become resistant to infection from unrelated viruses, fungi, and bacteria (40). Moreover, induction of PR proteins by abiotic elicitors generally results in acquired resistance to pathogens, with very few exceptions (36). Some well-known chemical inducers of PR proteins include polyacrylic acid, ethephon (which metabolizes into ethylene), aromatic compounds such as benzoic acid, salicylic acid and acetyl salicylic acid; amino acid derivatives; the antiviral agents 2-thiouracil and dioxohexahydrotriazine; the herbicide phosphinotricin; and barium and manganese salts (5, 118; J. F. Bol et al, unpublished results). It has been claimed that PR proteins are also induced by elicitor preparations extracted from fungal or bacterial cultures (65, 75). Injection of tobacco with a xylanase solution resulted in an induction of PR proteins, suggesting that released oligosaccharide fragments may act as signal molecules (70). On the other hand, spraying of tobacco plants with a glucan preparation from the mycelial walls of *Phytophthora megasperma* did not induce PR proteins or enzymes from the phenylpropanoid pathway (38).

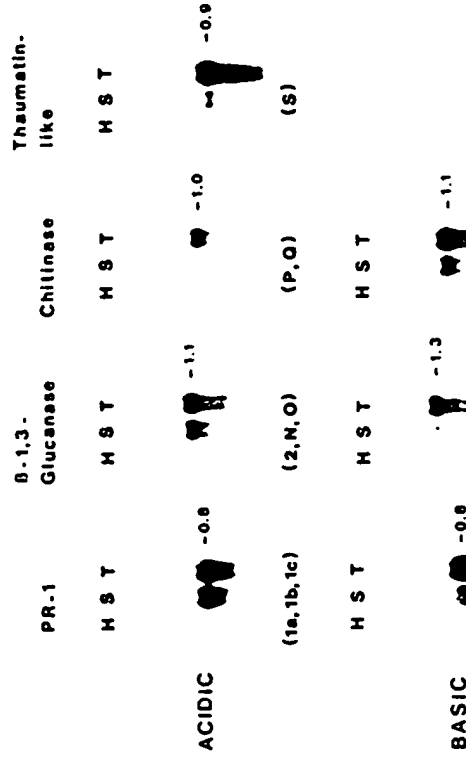


Figure 1 Induction of PR-mRNAs in Samsun tobacco (NN genotype) by treatment with salicylate or infection with TMV. Seven Northern blots were loaded with RNA from healthy plants (lanes H), plants sprayed with 5 mM salicylate (lanes S) and plants infected with TMV (lanes T). The blots were hybridized to  $^{32}$ P-labeled cDNA clones corresponding to the acidic (upper panels) and basic (lower panels) isoforms of PR-1 proteins.  $\beta$ -1,3-glucanases and chitinases, and the acidic isoform of PR-R/S. The estimated size of the mRNAs (kb) is indicated in the right margin of each blot. Reprinted, with permission of J. F. Bol et al (7).

Similarly, floatation of tobacco leaf discs on such an elicitor solution did not induce PR proteins (L. C. van Loon, personal communication).

Figure 1 shows seven Northern blots that reveal the accumulation of mRNAs encoding acidic and basic PR proteins in the leaves of healthy tobacco plants (lanes H), plants sprayed with a solution of 5 mM salicylate (lanes S), and plants inoculated with TMV (lane T). All mRNAs are present at a relatively low concentration in the leaves of healthy plants and are strongly induced after TMV infection. In addition, a number of mRNAs are induced by salicylate, notably those encoding acidic and basic PR-1 proteins, acidic glucanases, and basic chitinases. An analysis of a cDNA library made to poly(A)-RNA from TMV-infected tobacco yielded clones corresponding to several induced mRNAs that do not encode known PR proteins (51). Figure 2 is a Northern blot showing the induction by salicylate and TMV-infection of mRNA, called "A", encoding a 40-kd protein of unknown function, and of mRNA encoding a glycine-rich protein (GRP) (113).

For six of the mRNAs shown in Figures 1 and 2, accumulation reaches a maximum concentration about four days after inoculation with TMV in

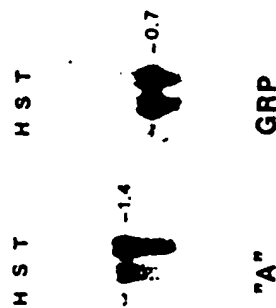


Figure 2 Induction of "A" and GRP mRNAs in Samsun tobacco (NN genotype) by salicylate treatment or TMV infection. Northern blots were loaded with RNA from healthy plants (lanes H), plants sprayed with 5 mM salicylate (lanes S) and plants infected with TMV (lanes T). The blots were hybridized to  $^{32}$ P-labeled "A" and GRP cDNA clones.

primary infected leaves. After eight days the mRNAs are detectable in virus-free, systemically induced leaves (51).

The expression of a number of PR genes is developmentally regulated. Chitinase,  $\beta$ -1,3-glucanase, and the corresponding mRNAs are present at high concentrations in roots of healthy plants (34, 103). Memelink et al (80) showed by Northern blot analysis that genes encoding basic PR-1 proteins—basic  $\beta$ -1,3-glucanases, and basic chitinases—are expressed at different levels in roots, stems, and flowers of healthy plants (Table 2). It has been reported that pricking leaves with needles moistened with ethephon, a treatment that causes necrotic lesion-like spots on the leaves, results in induction of genes encoding acidic PR proteins (116). Spraying the leaves with an ethephon solution did not induce the acidic PR proteins but efficiently "turned on" the genes encoding basic PRs and the non-PR proteins A and GRP (Table 2). Similarly, the genes encoding basic  $\beta$ -1,3-glucanases and basic chitinases are coordinately induced by exposure of plants to ethylene (11, 127).

The plus and minus signs in Table 2 represent crude estimates of the various mRNA levels. A more careful serological analysis revealed an organ-specific expression of several acidic PR proteins in different parts of the flowers of healthy tobacco (71). Acidic PR-1 proteins accumulate in sepals, whereas a glycosylated vacuolar form of acidic  $\beta$ -1,3-glucanase is found in pistils. Interestingly, within the class of acidic chitinases, the genes encoding PR proteins P and Q are differentially expressed in sepals (P and Q), pedicels (P only), anthers (low levels of P and Q), and ovaries (low levels of P only).

Stress-induced expression of tobacco GRP genes is leaf-specific and occurs in a light-dependent manner (80). The induction of acidic PR proteins by various chemicals in tobacco leaf material is light-dependent (3). The observation that PR genes differ in their response to salicylic acid or ethylene

Table 2 Accumulation of TMV-inducible mRNAs in tobacco Samsun (NN genotype).

mRNA	Virus infected leaves	Ethephon sprayed leaves	Healthy plants				
			leaves	roots	stems	flowers	fruits
Acidic PR PR-1	+	+	—	—	—	—	—
	Gluc.	+	—	—	—	—	—
	Chit.	+	—	—	—	—	—
	PR-S	+	—	—	—	—	—
Basic PR PR-1	+	+	+	+	+	+/—	—
	Gluc.	+	+	+	+/—	+/—	—
	Chit.	+	+	+	+/—	+/—	—
Non PR A	+	+	—	—	—	—	—
	GRP	+	—	—	—	—	—

\* Under conditions indicated by a plus sign the accumulation of mRNA is at least 1000-fold higher than under conditions indicated by a minus sign. For experimental details see Memelink et al (80).

and differ in their tissue-specific expression, indicates that the promoters of these genes are controlled by complex regulatory elements. The recent cloning of a number of PR genes is a first step toward characterizing these elements.

## PR-1 PROTEINS

cDNAs to tobacco mRNAs encoding acidic PR-1 proteins have been cloned by several groups (24, 27, 76, 92). Three types of mRNAs were identified that could be correlated, respectively, with PR-1a, -1b, and 1c by determining the partial amino acid sequences of these proteins (73, 91). The amino acid sequence identity between the three acidic proteins is over 90%. The 138-amino-acid-long mature proteins are derived from primary translation products by removal of an N-terminal signal peptide of 30 amino acids. The acidic PR-1 proteins from *Nicotiana tabacum* cvs. Samsun NN, Wisconsin-38, and Xanthi nc show minor sequence variations. Analysis of the cDNA clones revealed that alternate polyadenylation signals are used for each PR-1 mRNA.

Screening of a cDNA library made to poly(A)-RNA from TMV-infected tobacco yielded one clone encoding a basic isoform of the acidic PR-1 proteins (25). The amino acid sequence similarity between the acidic and basic PR-1 proteins is about 67%. The 19-kd basic protein is made with a C-terminal extension of 36 amino acids not present in the acid isoforms. A genomic clone has been identified encoding another basic PR-1 protein that contains a C-terminal extension of 18 amino acids (89).

Infection of tomato plants with either viroids, viruses, or *Cladosporium fulvum* results in the accumulation of a basic 14-kd protein (isoelectric point of 10.7), referred to as p14 or P1. Determination of the primary structure of p14 revealed a 60% amino acid sequence similarity with acid and basic PR-1 proteins of tobacco (73). Antisera to the tobacco and tomato proteins cross-react. Moreover, these antisera permitted the detection of serologically related inducible proteins in cowpea, potato, *Solanum demissum*, *Gomphrena globosa*, *Chenopodium amaranticolor*, maize, and barley (85, 130).

Because of their relative resistance to proteases, PR proteins are well adapted to survive in the intracellular space of the leaf or in vacuoles. Their long half-lives, estimated to be 40–70 hr, permit the accumulation of high levels of PR proteins. In viroid-infected tomato plants p14 was specifically degraded by a constitutive extracellular endoproteinase with a molecular weight of 37 kd (100); possibly, this enzyme plays a role in the turnover of PR proteins.

## GLUCANASES AND CHITINASES

### General Occurrence

The acidic or basic  $\beta$ -1,3-glucanases or chitinases that are coordinately induced in plants by ethylene, fungal elicitor, fungal or viral infection are all endo-hydrolases. Chitinases catalyze the hydrolysis of  $\beta$ -1,4-N-acetylglucosamine linkages of chitin polymers. Most of the purified plant chitinases tested have lysozyme activity, and some enzymes under optimal conditions hydrolyze bacterial peptidoglycan faster than chitin (12). Chitin is not present in healthy plants but may constitute a significant fraction of fungal cell walls, ranging from 0.3% in *Phytophthora* spp. to about 50% in *Fusarium* spp. The other major component in the cell wall of these microorganisms is  $\beta$ -1,3-glucan. Combinations of purified pea chitinase and  $\beta$ -1,3-glucanase inhibited growth of 15 out of 18 fungi tested, whereas the enzymes alone inhibited only one of these fungi (77). Possibly, these hydrolases play a role in the TMV-induced resistance of tobacco to subsequent infection by fungi and bacteria (40).

The combination of enzymes with  $\beta$ -1,3-glucanase and chitinase activity has been found so far in tobacco, potato, tomato, bean, soybean, pea, barley, and maize (12, 84). In addition, chitinases have been identified in ten other plant species including cucumber, melon, and carrot. Below we discuss some of the recently characterized systems.

### Tobacco

When tobacco leaf tissue was subcultured on an auxin-containing medium, up to 10% of the soluble protein fraction was found to consist of a 33-kd protein

that was identified as a basic  $\beta$ -1,3-glucanase (33). cDNA cloning indicated the induction under these conditions of at least three different mRNAs that encode highly homologous basic  $\beta$ -1,3-glucanases (104). A similar cDNA clone was used in Figure 1 to reveal the induction of these mRNAs by TMV-infection. The basic  $\beta$ -1,3-glucanase is produced as a 359-residue preproenzyme. In a first maturation step the N-terminal signal peptide of 21 residues is removed and the C-terminal sequence becomes glycosylated. Subsequently, a C-terminal sequence of 22 residues with the oligosaccharide side chain is lost (104). Most acidic tobacco  $\beta$ -1,3-glucanases are produced without this C-terminal extension. From an analysis of cDNA clones (H. J. M. Linthorst & J. F. Bol, unpublished observations) and partial protein sequence data, it can be concluded that at least five acidic  $\beta$ -1,3-glucanases with highly similar amino acid sequences are induced in tobacco leaves by TMV-infection. The amino acid sequence identity between basic and acidic  $\beta$ -1,3-glucanases is about 65%.

Analysis of cDNA and genomic clones revealed the sequence of a hormonally regulated  $\beta$ -glucanase from *Nicotiana plumbaginifolia* (28). This enzyme contains a C-terminal extension similar to that of the basic tobacco  $\beta$ -1,3-glucanase. The amino acid sequence identity between the two enzymes is 73%.

Sequencing of cDNA clones corresponding to basic chitinases (103) and acidic chitinases (50; H. J. M. Linthorst & J. F. Bol, unpublished observations) revealed that these PR proteins also are synthesized as precursors with a N-terminal signal peptide. The amino acid sequence identity between these two isoforms is about 65%. The higher molecular weight of the basic chitinases compared to the acidic chitinases (Table 1) is due to a N-terminal extension of 46 amino acids and a C-terminal extension of 6 amino acids (H. J. M. Linthorst & J. F. Bol, unpublished observations). Interestingly, on 40 residues the N-terminal of the basic chitinases show significant sequence similarities to the N-terminal domains of other chitin-binding proteins such as hevein, wheat-germ agglutinin, and rice and nettle lectins (12, 74, 109). The absence of this domain in acidic tobacco chitinases demonstrates that it is not involved in catalytic activity.

### Potato

After inoculation of leaves with zoospores of *Phytophthora infestans* or treatment of the leaves with an elicitor derived from *P. infestans* culture filtrate and administered through the petioles, potato plants start to accumulate massive amounts of at least nine major proteins having molecular weights in the range of 10–40 kd (60). Two proteins with molecular weights of 36 and 36.2 kd and isoelectric (pI) points of 9.6 and 9.8, respectively, were shown to be basic  $\beta$ -1,3-glucanases. These proteins were serologically related to  $\beta$ -1,3-

glucanases from bean (see below). Six other proteins with molecular weights between 32.6 and 38.7 kD proved to be basic chitinases with pI values above 7. All six chitinases showed strong cross-reactivity with a bean anti-chitinase antiserum.

cDNA was produced from a potato mRNA that encodes a basic chitinase with a predicted pI of 10.55. After removal of the signal peptide the molecular weight of the mature form is 32.4 kD (39). The enzyme shows considerable sequence similarity to the tobacco chitinases.

### Bean

Exposure of bean seedlings to ethylene results in the rapid and coordinate induction of a 32.5-kD chitinase and a 36-kD  $\beta$ -1,3-glucanase with pI values of 9.0 and 9.7, respectively (127). The use of cDNA clones to the mRNAs of chitinase (17) and  $\beta$ -1,3-glucanase (127) as probes, showed that induction of these PR proteins is regulated at the level of transcription. A similar rapid induction of chitinase mRNA was observed in bean hypocotyls in response to wounding or infection with *Colletotrichum lindemuthianum* and after treatment of bean protoplast cultures with fungal cell-wall elicitors (45). The cDNA sequence revealed that the mature chitinase of 301 residues is derived from a precursor with a 27-residue amino-terminal signal peptide. The amino acid sequence identity between the bean chitinase and the basic tobacco chitinase is 73% (103). The polypeptide translated in vitro from bean  $\beta$ -1,3-glucanase mRNA was 4 kD larger than the mature protein, which indicates that this enzyme also is derived from a precursor.

### Pea

Two basic chitinases and two basic  $\beta$ -1,3-glucanases have been identified in pea. The accumulation of these enzymes is differentially regulated. One chitinase (Ch2, mol wt 36.2 kD) and  $\beta$ -1,3-glucanase (G1, mol wt 33.5 kD) are produced in pea pods during the course of maturation. Another chitinase (Ch1, mol wt 33.1 kD) and  $\beta$ -1,3-glucanase (G2, mol wt 34.3 kD) are strongly induced in immature pea pods in response to various stress conditions such as inoculation with strains of *Fusarium solani* pathogenic or nonpathogenic to pea, wounding, or treatment with chitosan or ethylene (77). The developmentally regulated and stress-induced isozymes differ in a number of enzymatic properties, which indicates that they fulfill different functions.

### Cucumber

Infection of cucumber with tobacco necrosis virus or treatment with salicylate results in the induction of an acidic chitinase with molecular weight of 28

kD. The sequence derived from a cDNA clone shows that the protein of 267 residues matures from a precursor with a N-terminal signal peptide of 25 amino acids (81). The amino acid sequences of the known chitinases from tobacco, potato, bean, and barley are closely related to each other, but no detectable similarity is observable between these enzymes and the cucumber chitinase. However, this latter enzyme shows striking homology to the partial amino acid sequence deduced from a lysozyme/chitinase of *Parthenococcus quinquefolia*.

### Barley

Two classes of basic  $\beta$ -glucan endohydrolases in barley have been characterized at the molecular level, each represented by two isozymes. The first class contains  $\beta$ -1,3-1,4-glucanases (isozymes EI and EII) that are primarily responsible for the degradation of cell walls of the nonliving starchy endosperm during germination of barley grains. The second class consists of  $\beta$ -1,3-glucanases (isozymes GI and GII) that also accumulate to relatively high concentrations in barley grains although their substrate in germinating barley is very limited. The observation that after infection with *Erysiphe graminis*  $\beta$ -1,3-glucanase mRNA accumulates to high concentrations in resistant plants but not in susceptible plants, indicates that GI and GII are involved in defense reactions (48). cDNA to EI (35) and GII (48) has been cloned and sequenced. Both proteins are basic polypeptides of 306 amino acids, derived from their respective precursors by removal of the N-terminal signal peptide. The amino acid sequence identity between EI and GII is about 50% and a similar homology is observed with the basic tobacco  $\beta$ -1,3-glucanase.

Another chitinase, a 36-kD endochitinase, is secreted from cells in the aleurone layer during seed germination. cDNA cloning of the mRNA of this chitinase revealed an extensive homology to the tobacco and bean chitinase and also to a 28-kD endochitinase present in barley flour (110).

### Maize

Eight proteins, all having molecular weights in the range of 14.2–34.5 kD, were found to accumulate in maize leaves after treatment with mercuric chloride or infection with bromo mosaic virus. These were named PRm 1 to 8 (84), and could be divided into at least three classes. PRm 2 (16.5 kD) was serologically related to the tobacco PR-1b protein. Four of these proteins (PRm 3, 4, 5, and 7) were endochitinases. PRm 3 and 4 both have a molecular weight of 25 kD and are serologically related. PRm 5 and 7 have molecular weights of 29 kD and 34.5 kD, respectively; they are serologically related and cross-reacted with an antiserum to the tobacco chitinase PR-P, but are not serologically related to PRm 3 and 4. Finally, PRm 6a (32 kD) and PRm 6b (30.5 kD) appeared to be serologically related  $\beta$ -1,3-glucanases.

## THAUMATIN-LIKE PROTEINS

Pierpoint et al (95) found two closely related acidic 24-kd proteins, named R-minor and R-major, that occur in a 2:3 ratio in TMV-infected *Nicotiana tabacum* L. cv Xanthi nc tobacco. These proteins probably correspond to PR-R and PR-S, respectively, in Fritig's (38) nomenclature. cDNA has been cloned, respectively, to the mRNAs of PR-R (23) and PR-S (90). Salt stress of tobacco cells induces two related basic 24-kd proteins, named osmotin I and II, that occur in an approximate ratio of 2:3 (105). Determination of the N-terminal amino acid sequence revealed a 62% homology with PR-R/S. cDNA cloning of a salt-induced basic 24-kd tomato protein (NP24) revealed that this protein is over 90% homologous to the tobacco osmotin (59). The acidic and basic tobacco proteins and the tomato NP24 all show approximately 60% amino acid sequence homology to the sweet-tasting protein thaumatin, which accumulates in the fruits of *Thaumatococcus daniellii*, and to a bifunctional amylase/protease-inhibitor from maize (23, 97). The acidic and basic tobacco proteins do not taste sweet. The basic thaumatin is synthesized as a precursor with an N-terminal signal peptide and a C-terminal extension of six amino acids that are removed during maturation. The tobacco and tomato thaumatin-like proteins also have signal peptides. The basic tomato protein (and possibly also the basic tobacco protein) has a C-terminal extension of 19 residues that is absent in the acidic tobacco protein.

## TARGETING OF PR PROTEINS

In tobacco, acidic and basic PR proteins appear to be strictly compartmentalized. The former accumulate predominantly in the extracellular space of the leaf, i.e. the apoplast (88), which has been analyzed in most detail by cell fractionation techniques for acidic PR-1 proteins (86) and  $\beta$ -1,3-glucanases (112). The extracellular location of acidic PR-1 was verified by immunofluorescence microscopy and immunogold-labeling techniques (19, 30, 52). On the other hand, according to available evidence, all basic tobacco PR proteins are sequestered in vacuoles. Immunocytochemical detection revealed that osmotin is concentrated in dense inclusion bodies within the vacuole. Using cell fractionation techniques, the basic tobacco  $\beta$ -1,3-glucanase (112) and chitinase (unpublished observation quoted in 55) were also shown to accumulate in vacuoles. In tobacco, the extracellular acidic PR proteins may serve as a first line of defense against invading pathogens, with the vacuolar basic proteins stored in the vacuole as a second line of defense (9).

Although the basic PR-1 protein in tobacco is not secreted in detectable amounts, its equivalent basic p14 protein in tomato is detectable in relatively

large quantities in the intercellular fluid. By using immunogold-labeling techniques, p14 was detectable in the intercellular spaces as well as in association with inclusion bodies within the vacuoles of viroid-infected tomato plants, which suggests a dual localization (124). The basic thaumatin-like NP24 protein of tomato was found to accumulate intracellularly (59).

In potato all basic chitinases and  $\beta$ -1,3-glucanases were predominantly found in the intercellular fluid, which suggests that they are secreted after fungal infection or elicitor treatment (60). Similarly, the basic chitinases and  $\beta$ -1,3-glucanases of barley are secreted by cells in the barley aleurone layer (48, 110). In contrast, cell fractionation and immunocytochemical studies on ethylene-treated bean leaves revealed that all of the basic bean chitinase and most of the basic bean  $\beta$ -1,3-glucanase accumulated in vacuoles (14, 78, 127). Only a small amount of the bean  $\beta$ -1,3-glucanase, but no chitinase, is secreted into the apoplast. Like the acidic tobacco chitinase, the acidic chitinase of cucumber is secreted almost exclusively into the intercellular space of infected leaves (13).

All known PR proteins are synthesized as precursors with an N-terminal signal peptide. Little homology is observed between these signal sequences, except for their hydrophobic nature and the presence of charged residues at the preferred positions (129). Possibly, they function only in the translocation of the PR polypeptide chains across the membrane of the endoplasmic reticulum, and signals in the mature proteins determine whether they are transported from the Golgi complex to the vacuoles or are secreted through the plasma membrane. The current view is that a positive signal near the N-terminus of a mature protein specifies its routing to the vacuoles, whereas the absence of such a signal results in a default pathway leading to secretion (for references see 125, 126). This would mean that the C-terminal extension present in the vacuolar basic PR proteins of tobacco, but absent in the extracellular acidic isoforms, has no role in the targeting of these proteins. Similarly, the presence of an N-terminal sequence of 46 amino acids in the basic tobacco chitinase, which is absent in the acidic tobacco chitinase, is not responsible per se for the vacuolar localization of the basic isoform because a similar N-terminal sequence is present in the extracellular basic potato chitinase. Possibly, a modulation of a putative signal near the N-terminus of mature PR proteins determines whether the protein is transported to the vacuoles, is secreted to the apoplast, or whether both pathways are followed. The identification of signals involved in targeting of PR proteins requires further experimentation.

Only two examples of glycosylation of PR proteins are known. The C-terminal extension of the basic tobacco  $\beta$ -1,3-glucanase is glycosylated but the oligosaccharide side chain is not present in the mature protein (104). Moreover, a glycosylated vacuolar form of acidic  $\beta$ -1,3-glucanase has been



identified in flower parts of healthy tobacco (71). Available evidence indicates that glycosylation contributes to the stability of a protein but has no role in its targeting (126). Nevertheless, it is interesting to note the similarity in processing of basic tobacco  $\beta$ -1,3-glucanase and the vacuolar wheat germ agglutinin (WGA). Also, WGA is glycosylated at a C-terminal extension of 15 amino acid concomitantly with the removal of a signal peptide and this extension is cleaved off during transport to the vacuole (96).

## OTHER VIRUS-INDUCED PROTEINS

### *Biosynthesis of Aromatic Compounds*

The hypersensitive response of plants to pathogens is accompanied by increased activity of enzymes involved in the biosynthesis of aromatic compounds based on the phenylpropane skeleton. These compounds include the cell wall structural polymer lignin, flavonoid pigments and UV protectants, furanocoumarin and isoflavonoid phytoalexins, and wound protectant hydroxycinnamic acid esters. The general pathway in phenylpropanoid metabolism involves the conversion of phenylalanine to 4-coumaroyl-CoA in three steps catalyzed by phenylalanine-ammonia-lyase (PAL), cinnamate-4-hydroxylase (C4H), and 4-coumaroyl-CoA ligase (4CL). From this route different branches lead to the synthesis of furanocoumarins, (iso)flavonoids, stilbenoids, suberin, lignin, and other phenolics. Chalcone synthase (CHS) and chalcone flavanone isomerase (CHI) are the key enzymes in the branch toward (iso)flavonoid synthesis. Cinnamoyl-CoA-reductase and cinnamoyl-alcohol-dehydrogenase (CAD) generate the lignin precursor alcohols 4-coumaryl, coniferyl, and sinapyl, which are polymerized by peroxidases to give the three-dimensional lignin network. Recently, the following genes involved in the biosynthesis of these aromatic compounds have been cloned and characterized: PAL from bean and parsley (26, 69), 4CL from parsley (72), CHS from parsley, soybean, *Antirrhinum majus*, zea mays, arabidopsis, petunia and bean (32, 46, 131), CHI from petunia (122), and CAD from bean (128). Most of these genes correspond to multigene families that are differentially regulated. The induction of PAL, 4CL, CHS, CHI, and CAD genes by infection, elicitor, and ethylene has been demonstrated (20, 31, 64, 117). Initial experiments on the characterization of regulatory elements in the promoter sequences of these genes have been performed (29, 63).

### *Peroxidases*

At least 12 peroxidase isozymes divisible into three subgroups have been identified in tobacco (61). The function of the vacuolar cationic peroxidases (pI, 8.1–11) is as yet unclear. The moderate anionic (pI, 4.5–6.5) and the anionic (pI, 3.5–4.0) peroxidases are cell-wall associated and are thought to

function in suberization and lignification, respectively. After TMV infection, two of the moderate cationic peroxidases are formed in inoculated and systemically induced tobacco leaves (62, 117). The molecular cloning of an anionic lignin-forming peroxidase from tobacco (61) and suberization-associated anionic peroxidases from potato (99) and tomato (98) has been reported. The enzymes show a 40–50% sequence homology to horseradish and turnip peroxidase.

### *Hydroxyproline-Rich and Glycine-Rich Proteins*

Hydroxyproline-rich glycoproteins (HRGP, extensin) make up 1–10% of the wall matrix. Intermolecular cross-linking of extensin monomers by peroxidases results in a network that serves as a matrix for the linkage of polysaccharides (pectin, cellulose) or polyphenols (lignin, suberin) (22). Synthesis of HRGP mRNA is induced by ethylene, fungal infection, elicitors (31, 64), and by infection of tobacco with TMV (79).

Certain plants contain little if any HRGP; instead their cell walls are rich in glycine. A wound-inducible glycine-rich protein (GRP) from petunia was suggested to be a cell wall component (21). Association of a wound-inducible GRP with the vascular system of bean has been confirmed by immunocytochemical techniques (57). Like the TMV-inducible tobacco GRP shown in Figure 2, the petunia and bean GRP are synthesized with N-terminal signal peptides. Interestingly, the sequences of the signal peptides of the tobacco GRP and a light-induced GRP from *Chenopodium rubrum* are highly similar (54, 113). Induction of GRPs by water stress or abscisic acid has been observed in maize and rice, but these are probably cytosolic proteins (44, 83).

### *Active Defense Proteins*

Several other inducible plant proteins besides the group of PR proteins are involved in defense against pathogenic attack. Low molecular weight proteinase inhibitors may constitute up to 5% of the protein in seeds and tubers of Gramineae, Leguminosae and Solanaceae. Their systemic induction in tomato and potato leaves by wounding is thought to be mediated by oligosaccharides released from plant-cell walls by an endopolygalacturonase (101; Ryan, this volume). These proteins inhibit serine endopeptidases that are found in both animals and microorganisms but only rarely in plants. Because of this specificity they are thought to be involved in defending the plant against herbivorous insects.

Pathogens and various stress conditions trigger the synthesis of thionins in the leaves of barley, maize, wheat, oats, and several dicotyledonous plants. These proteins are incorporated in the cell wall and display significant antifungal activity in *in vitro* assays (4).

TMV-infected protoplasts from hypersensitive tobacco cultivars have been

reported to secrete an inhibitor of virus replication (IVR) into the medium. Virus replication was inhibited in protoplasts and leaf disks and also in intact leaves when IVR was applied by cut stems or by spray. IVR was identified as a 23-kD protein that is not only released by infected protoplasts but also accumulates in the intercellular space of TMV-infected tobacco (108). IVR is not serologically related to AVF, a putative antiviral factor consisting of a 22-kD phosphoglycoprotein (82). A possible serological cross-reaction of IVR with well-characterized tobacco PR proteins has not yet been analyzed.

### Other Inducible Proteins

Differential screening of a cDNA library from cultured parsley cells yielded 18 independent cDNA families of elicitor-induced mRNAs (106). The structure and sequence of a gene corresponding to one of these classes of parsley mRNAs, named PR1, has been deduced (107). This parsley PR1 gene does not show sequence similarity to other known PR genes. Moreover, the 16.5-kD parsley PR1 protein (isoelectric point of 4.7) is synthesized without a signal peptide, suggesting that it accumulates in the cytosol. The PR1 mRNA is induced by fungal infection of parsley.

In addition to p14, a 69-kD PR protein, referred to as P-69, is induced in tomato by citrus exocortis viroid infection. Recently, it was shown that P-69 is an alkaline endoproteinase (pI, 8.5 to 9.0) with optimal activity at pH 8.5 (123). It has been suggested that P-69 is involved in degenerative processes occurring in the chloroplasts of infected plants.

Treatment of *Nicotiana plumbaginifolia* with ethylene or salicylic acid, or infection of this plant with *Pseudomonas syringae* results in the induction of the enzyme manganese superoxide dismutase. cDNA cloning revealed that the enzyme is synthesized with an N-terminal leader sequence resembling a transit peptide for mitochondrial targeting. The function of the enzyme could be the dismutation of superoxide radicals generated by the enhanced respiratory oxidation of sugars when stress is applied to the plant (15).

## REGULATION OF PR GENE EXPRESSION

### PR-1 and GRP Genes

Among the salicylate-inducible tobacco genes shown in Figures 1 and 2, the PR-1 and GRP genes have been characterized in most detail. The gene families encoding acidic PR-1, basic PR-1, and GRP from Samsun NN tobacco each consist of about eight members. After TMV infection, three genes encoding acidic PR-1 proteins (PR-1a, -1b and -1c) are expressed and at least two genes encoding basic PR-1 proteins and four genes encoding GRP (25, 89, 92; H. J. M. Linthorst & J. F. Bol, unpublished observation). The PR-1a and -1c genes are derived from *Nicotiana glauca* whereas the PR-1b

gene is derived from the other parent of Samsun NN tobacco, *N. tomentosiformis* (41). The sequence of the PR-1a gene and several pseudo genes from the PR-1 family has been reported (25, 87, 91, 93).

One gene encoding a basic PR-1 protein has been cloned (89). There is limited homology with the PR-1a gene throughout the 5'-flanking sequences with no obvious conserved region. The cloned gene encoding basic PR-1 is not coordinately expressed with genes encoding acidic PR-1, in contrast to the gene corresponding to the basic PR-1 cDNA clone used to probe the blot of Figure 1. This suggests that genes encoding basic PR-1 proteins are regulated differently.

Two genomic GRP clones have been sequenced (113). One gene, named GRP-8, contains an intron of 555 bp and is expressed after TMV infection, whereas the other gene (GRP-4) is not expressed after TMV infection and contains an intron of 1954 bp. A defective splice donor site was thought to be responsible for the "silent" nature of the GRP-4 gene.

The upstream sequences of the PR-1a and GRP genes contain a number of direct and inverted repeats that could have regulatory functions. Most prominent is a 64-bp inverted repeat in the GRP upstream sequence that is almost identical to a 64-bp inverted repeat found upstream of the rubisco small subunit gene of tobacco (113). Upstream sequences of the PR-1a and GRP-8 genes have been fused to reporter genes, and these constructs were used to transform tobacco. In the transgenic plants, 689 bp of the PR-1a upstream sequence and 645 bp of the GRP upstream sequence were sufficient to induce expression of the reporter gene when the plant was sprayed with 5 mM salicylate or infected with TMV (M. D. Van de Rhee, J. A. L. van Kan, M. T. González-Jaén & J. F. Bol, unpublished observations). Deletion studies revealed that sequences located between positions -689 and -643 in the PR-1a promoter and positions -645 and -400 in the GRP-8 promoter were required for induction of reporter-gene expression by the two treatments. The only detectable similarity between the two promoter regions was a small sequence resembling the SV40 core enhancer sequence. Circumstantial evidence suggested that such an enhancer sequence could work in concert with salicylate-responsive elements located near the TATA-box (M. D. Van de Rhee, J. A. L. van Kan, M. T. González-Jaén & J. F. Bol, unpublished observation). Similar fusions of the promoter region of the gene for the bean cell-wall glycine-rich protein with a reporter gene revealed that a 494-bp upstream sequence was sufficient for correct tissue-specific expression and wound-inducibility of the gene (58).

### Glucanase and Chitinase Genes

The genome of Samsun NN tobacco contains about eight genes encoding acidic  $\beta$ -1,3-glucanases and a similar number encoding basic  $\beta$ -1,3-

cell walls constitute the signal(s) that trigger(s) the induction of defense genes at a distance from the site of infection (see Ryan, this volume).

## CONCLUDING REMARKS

The host proteins induced by the hypersensitive response of plants to infection with viruses or other pathogens appear to be involved in at least three types of defense reactions: (a) a direct attack of the invading pathogen (e.g. hydrolytic enzymes), (b) a localization of the pathogen at the site of infection (e.g. enzymes involved in lignification), and (c) an adaptation of the host metabolism to the stress condition (e.g. superoxide-dismutase). The genes encoding these inducible proteins possess interesting features, each pointing to future areas of productive research. Constitutive expression of these genes in genetically engineered plants may shed light on the role of the corresponding proteins in defense reactions and the function of their tissue-specific expression in mature or senescing plants. This insight could be used to increase the natural resistance of economically important crops.

A study of promoter regions will reveal the similarities and differences in the cis-acting elements involved in the induction of these genes by infection, elicitors, ethylene, other plant hormones, wounding, abiotic inducers, etc. Identification of the transacting factors binding to these cis-acting elements will help unravel the diversity of the induction pathways and may permit the linkage of these intracellular pathways with the intercellular signal molecules responsible for systemic induction of defense genes. Finally, the expression of mutated or chimeric proteins in plants will permit an analysis of polypeptide sequences that determine whether the protein is targeted to the vacuoles or is destined to follow the intercellular secretion pathway. Thus, studies on plant defense genes will also contribute to our general understanding of the molecular biology of plant cells.

## Literature Cited

- Ahl, P., Gianazzi, S. 1982. *b*-protein as a constitutive component in highly (TMV) resistant interspecific hybrids of *Nicotiana glauca* × *Nicotiana debneyi*. *Plant Sci. Lett.* 26:173-81.
- Antoniw, J. F., Ritter, C. E., Pierpoint, W. S., Van Loon, L. C. 1980. Comparison of three pathogenesis-related proteins from plants of two cultivars of tobacco infected with TMV. *J. Gen. Virol.* 47:79-87.
- Asselin, A., Grenier, J., Cote, F. 1985. Light-influenced extra cellular accumulation of *b* (pathogenesis-related) proteins in *Nicotiana glauca* green tissue induced by various chemicals or prolonged flotation on water. *Can. J. Bot.* 63: 1276-83.
- Bohmann, H., Clausen, S., Behnke, S., Giese, H., Hiltner, C., et al. 1988. Leaf-specific thionins of barley—a novel class of cell wall proteins toxic to plant-pathogenic fungi and possibly involved in the defence mechanism of plants. *EMBO J.* 7:1559-65.
- Bol, J. P. 1988. Structure and expression of plant genes encoding pathogenesis-related proteins. In *Plant Gene Research: Temporal and Spatial Regulation of Plant Genes*, ed. D. P. S. Verma,
- R. D. Goldberg, pp. 201-21. *Wien/New York: Springer-Verlag*.
- Bol, J. F., Linthorst, H. J. M., Neuwissen, R. L. J. 1990. Molecular responses of tobacco to virus infection. In *Biochemistry and Molecular Biology of Plant Pathogen Interactions*, ed. C. J. Smith. Oxford: Oxford Univ. Press. In press.
- Bol, J. F., Van de Rhee, M. D., Van Kan, J. A. L., González Jacin, M. T., Linthorst, H. J. M. 1989. Characterization of two virus-inducible plant promoters. See Ref. 74a, pp. 169-74.
- Bol, J. F., Van Kan, J. A. L. 1988. The synthesis and possible functions of virus-induced proteins in plants. *Microbiol. Sci.* 5:47-52.
- Bol, J. F., Van Kan, J. A. L., Cornelissen, B. J. C. 1988. Plant defense genes induced by virus infection. In *Molecular Biology of Plant-Pathogen Interactions*, ed. B. Staskiewicz, P. Ahlquist, O. Yoder, pp. 247-54. New York: Liss.
- Bol, J. F., Van Rossum, C. M. A., Cornelissen, B. J. C., Linthorst, H. J. M. 1990. Induction of host genes by the hypersensitive response of tobacco to virus infection. In *Viral Genes and Plant Pathogenesis*, ed. J. G. Shaw, T. P. Pirone. *Wien/New York: Springer-Verlag*. In press.
- Boller, T. 1985. Induction of hydrolases as a defense reaction against pathogens. In *Cellular and Molecular Biology of Plant Stress*, ed. J. L. Key, T. Kosuge, pp. 247-62. New York: Liss.
- Boller, T. 1988. Ethylene and the regulation of antifungal hydrolases in plants. In *Oxford Surveys of Plant Molecular and Cell Biology*, ed. B. J. Mullin, 5:145-75. Oxford: Oxford Univ. Press.
- Boller, T., Meittraux, J. P. 1988. Extracellular localization of chitinase in cucumber. *Physiol. Mol. Plant Pathol.* 33:11-16.
- Boller, T., Vögeli, U. 1984. Vacuolar localization of ethylene-induced chitinase in bean leaves. *Plant Physiol.* 74:442-44.
- Bowler, C., Alliotte, T., De Loose, M., Van Montagu, M., Inzé, D. 1989. The induction of manganese superoxide dismutase in response to stress in *Nicotiana glauca*. *EMBO J.* 8:31-38.
- Broglic, K. E., Biddle, P., Cressman, R., Broglic, R. 1989. Functional analysis of DNA sequences responsible for ethylene regulation of a bean chitinase gene in transgenic tobacco. *Plant Cell* 1:599-607.
- Broglic, K. E., Gaynor, J. J., Broglic, R. M. 1986. Ethylene-regulated gene
- expression: molecular cloning of the genes encoding an endochitinase from *Phaseolus vulgaris*. *Proc. Natl. Acad. Sci. USA* 83:6820-24.
- Carr, J. P., Beachy, R. N., Klessig, D. F. 1989. Are the PR1 proteins of tobacco involved in genetically engineered resistance to TMV? *Virology* 169:470-73.
- Carr, J. P., Dixon, D. C., Nikolau, B. J., Voelkeling, K. V., Klessig, D. F. 1987. Synthesis and localization of pathogenesis-related proteins in tobacco. *Mol. Cell. Biol.* 7:1580-83.
- Collinge, D. B., Slusarenko, A. J. 1987. Plant gene expression in response to pathogens. *Plant Mol. Biol.* 9:389-410.
- Condit, C. M., Meagher, R. B. 1987. Expression of a gene encoding a glycin-rich protein in petunia. *Mol. Cell. Biol.* 7:4273-79.
- Cooper, J. B., Chen, J. A., Van Holst, G.-J., Vamer, J. E. 1987. Hydroxyproline-rich glycoproteins of plant cell walls. *Trends Biol. Sci.* 12:24-27.
- Cornelissen, B. J. C., Hooft van Huysduynen, R. A. M., Bol, J. F. 1986. A tobacco mosaic virus-induced tobacco protein is homologous to the sweet-tasting protein thaumatin. *Nature* 321:531-32.
- Cornelissen, B. J. C., Hooft van Huysduynen, R. A. M., Van Loon, L. C., Bol, J. F. 1986. Molecular characterization of messenger RNAs for "pathogenesis-related" proteins 1a, 1b and 1c, induced by TMV infection of tobacco. *EMBO J.* 5:37-40.
- Cornelissen, B. J. C., Horowitz, J., Van Kan, J. A. L., Goldberg, R. B., Bol, J. F. 1987. Structure of tobacco genes encoding pathogenesis-related proteins from the PR-1 group. *Nucleic Acids Res.* 15:6799-11.
- Cramer, C. L., Edwards, K., Dron, M., Liang, X., Dildine, S., et al. 1989. Phenylalanine ammonia-lyase gene organization and structure. *Plant Mol. Biol.* 12:367-83.
- Cutt, J. R., Dixon, D. C., Carr, J. P., Klessig, D. F. 1988. Isolation and nucleotide sequence of cDNA clones for the pathogenesis-related proteins PR1a, PR1b and PR1c of *Nicotiana tabacum* cv. Xanthi nc induced by TMV infection. *Nucleic Acids Res.* 16:9861.
- De Lencastre, M., Alliotte, T., Chelyen, G., Gienel, C., Chien, J., et al. 1989. Primary structure of a hormonal regulated  $\beta$ -glucanase of *Nicotiana glauca*. *Gene* 70:13-23.
- Dron, M., Clouse, S. D., Dixon, R. A.

- Lawton, M. A., Lamb, C. J. 1988. Glutathione and fungal elicitor regulation of a plant defense gene promoter in electroporated protoplasts. *Proc. Natl. Acad. Sci. USA* 85:6738-42.
30. Dumas, E., Lherminier, J., Gianinazzi, S., White, R. F., Antoniow, J. F. 1988. Immunocytochemical location of pathogenesis-related b1 protein induced in tobacco mosaic virus-infected or polyacrylic acid-treated tobacco plants. *J. Gen. Virol.* 69:2687-94.
31. Ecker, J. R., Davis, R. W. 1987. Plant defense genes are regulated by ethylene. *Proc. Natl. Acad. Sci. USA* 84:5202-6.
32. Feinbaum, R. L., Ausubel, F. M. 1988. Transcriptional regulation of the *Arabidopsis thaliana* chalcone synthase gene. *Mol. Cell. Biol.* 8:1985-92.
33. Felix, G., Meins, F. 1985. Purification, immunoassay and characterization of an abundant, cytokinin-regulated polypeptide in cultured tobacco tissues. Evidence the protein is a  $\beta$ -1,3-glucanase. *Planta* 164:423-28.
34. Felix, G., Meins, F. 1986. Developmental and hormonal regulation of  $\beta$ -1,3-glucanase in tobacco. *Planta* 167:206-11.
35. Fincher, G. B., Lock, P. A., Morgan, M. M., Lingelbach, K., Wittenhall, R. E. H., et al. 1986. Primary structure of the (1,3)-4-D-glucan 4-glucosyltransferase from barley aleurone. *Proc. Natl. Acad. Sci. USA* 83:2081-85.
36. Fraser, R. S. S., Clay, C. M. 1983. Pathogenesis-related proteins and acquired systemic resistance: causal relationship or separate effects? *Neil. J. Plant Pathol.* 89:283-92.
37. Fritig, B., Kauffmann, S., Dumas, B., Geoffroy, P., Kopp, M., et al. 1987. Mechanism of the hypersensitive reaction of plants. In *Plant Resistance to Viruses*, ed. D. Evered, S. Hammett, Ciba Found. Symp. 133:92-108. Chichester: Wiley & Sons.
38. Fritig, B., Rueter, J., Kauffmann, S., Stintzi, A., Geoffroy, P., et al. 1989. See Ref. 74a, pp. 161-68.
39. Gaynor, J. J. 1988. Primary structure of an endochitinase mRNA from *Solanum tuberosum*. *Nucleic Acids Res.* 16:5210.
40. Gianinazzi, S. 1983. Genetic and molecular aspects of resistance induced by infection or chemicals. In *Plant Microbe Interactions: Molecular and Genetic Perspectives*, ed. E. W. Nester, T. Kossuge, 1:321-42. New York: Macmillan.
41. Gianinazzi, S., Ahi, P. 1983. The genetic and molecular basis of  $\beta$ -proteins in the genus *Nicotiana*. *Neil. J. Plant Pathol.* 89:275-81.
42. Gianinazzi, S., Martin, C., Vallee, J. C. 1970. Hypersensibilité aux virus, températures et protéines solubles chez le *Nicotiana glauca* L. Apparition de nouvelles macromolécules lors de la répression de la synthèse virale. *C. R. Acad. Sci. Paris D* 270:2383-86.
43. Gianinazzi, S., Pratt, H. M., Shewry, P. R., Miflin, B. J. 1977. Partial purification and preliminary characterization of soluble proteins specific to virus infected tobacco plants. *J. Gen. Virol.* 34:345-51.
44. Gómez, J., Sánchez-Martínez, D., Stiefel, V., Rigau, J., Puigdomènech, P., et al. 1988. A gene induced by the plant hormone abscisic acid in response to water stress encodes a glycine-rich protein. *Nature* 334:262-64.
45. Hedrick, S. A., Bell, J. N., Boller, T., Lamb, C. J. 1988. Chitinase cDNA cloning and mRNA induction by fungal elicitor, wounding and infection. *Plant Physiol.* 86:182-86.
46. Herrmann, A., Schulz, W., Hahlbrock, K. 1988. Two alleles of the single-copy chalcone synthase gene in parsley differ by a transposon-like element. *Mol. Gen. Genet.* 212:93-98.
47. Hilder, V. A., Gatehouse, A. M. R., Sheermann, S. E., Barker, R. F., Boulter, D. 1987. A novel mechanism of insect resistance engineered into tobacco. *Nature* 300:160-63.
48. Hoi, P. B., Hartman, D., Murrice, N., A., Duan, D. N. P., Fincher, G. B. 1989. Purification of (1,3)- $\beta$ -glucan endohydrolase isozyme II from germinated barley and determination of its primary structure from a cDNA clone. *Plant Mol. Biol.* 13:31-42.
49. Hooft van Huijsduijn, R. A. M., Abbas, S. W., De Rijk, R. H., Bol, J. F. 1986. Induction by salicylic acid of pathogenesis-related proteins and resistance to alfalfa mosaic virus infection in various plant species. *J. Gen. Virol.* 67:2135-43.
50. Hooft van Huijsduijn, R. A. M., Kauffmann, S., Brederode, F. Th., Cornelissen, B. J. C., Legrand, M., et al. 1987. Homology between chitinases that are induced by TMV infection of tobacco. *Plant Mol. Biol.* 9:411-20.
51. Hooft van Huijsduijn, R. A. M., Van Loon, L. C., Bol, J. F. 1986. cDNA cloning of six mRNAs induced by TMV infection of tobacco and a characterization of their translation products. *EMBO J.* 5:2057-61.
52. Hosokawa, D., Olashi, Y. 1988. Immunological localization of pathogenesis-related proteins secreted into the intercellular spaces of salicylate-treated tobacco leaves. *Plant Cell Physiol.* 29:88-128.
53. Janet, E., Fritig, B. 1986. Purification and characterization of 8 of the pathogenesis-related proteins in tobacco leaves reacting hypersensitively to tobacco mosaic virus. *Plant Mol. Biol.* 6:69-80.
54. Kaldenhoff, R., Richter, G. 1989. Sequence of cDNA for a novel light-induced glycine-rich protein. *Nucleic Acids Res.* 17:2853.
55. Kauffmann, S. 1988. Les protéines PR (pathogenesis-related) du tabac: des protéines impliquées dans les réactions de défense aux agents phytopathogènes. *Isolément, propriétés sérologiques et activités biologiques*. PhD thesis. Univ. Louis Pasteur, Strasbourg. 85 pp.
56. Kauffmann, S., Legrand, M., Geoffroy, P., Fritig, B. 1987. Biological function of "pathogenesis-related" proteins. Four PR-proteins of tobacco have 1,3- $\beta$ -glucanase activity. *EMBO J.* 6:3209-12.
57. Keller, B., Sauer, N., Lamb, C. J. 1988. Glycine-rich cell wall proteins in bean: gene structure and association of the protein with the vascular system. *EMBO J.* 7:3625-33.
58. Keller, B., Schmid, J., Lamb, C. J. 1989. Vascular expression of a bean cell wall glycine-rich protein- $\beta$ -glucuronidase gene fusion in transgenic tobacco. *EMBO J.* 8:1319-14.
59. King, G. J., Turner, V. A., Hussey, V. E., Wurtele, E. S., Lee, S. M. 1988. Isolation and characterization of a tomato cDNA clone which codes for a salicylate-induced protein. *Plant Mol. Biol.* 10:401-12.
60. Kombrink, E., Schröder, M., Hahlbrock, K. 1988. Several "pathogenesis-related" proteins in potato are 1,3- $\beta$ -glucanases and chitinases. *Proc. Natl. Acad. Sci. USA* 85:782-86.
61. Lagrimini, L. M., Burkhardt, W., Moyer, M., Rothstein, S. 1987. Molecular cloning of complementary DNA encoding the lignin-forming peroxidase from tobacco: molecular analysis and tissue-specific expression. *Proc. Natl. Acad. Sci. USA* 84:7542-46.
62. Lagrimini, L. M., Rothstein, S. 1987. Tissue specificity of tobacco peroxidase isozymes and their induction by wounding and tobacco mosaic virus infection. *Plant Physiol.* 84:438-42.
63. Lawton, M. A., Clouse, S. D., Lamb, C. J. 1990. Glutathione-elicited changes in chromatin configuration in the promoter of the plant defense gene chalcone synthase. *EMBO J.* In press.
64. Lawton, M. A., Lamb, C. J. 1987. Transcriptional activation of plant defense genes by fungal elicitor, wounding and infection. *Mol. Cell. Biol.* 7:335-41.
65. Leach, J. E., Sherwood, J., Fulton, R. W., Sequeira, L. 1983. Comparison of soluble proteins associated with disease resistance induced by bacterial lipopolysaccharide and by viral necrosis. *Physiol. Plant Pathol.* 23:377-85.
66. Legrand, M., Kauffmann, S., Geoffroy, P., Fritig, B. 1987. Biological function of "pathogenesis-related" proteins: four tobacco PR-proteins are chitinases. *Proc. Natl. Acad. Sci. USA* 84:6750-54.
67. Linthorst, H. J. M., Cornelissen, B. J. C., Van Kan, J. A. L., Van de Rhee, M., Meuwissen, R. L. J., et al. 1990. Induction of plant genes by compatible and incompatible virus-plant interactions. In *Recognition and Response in Plant-Virus Interactions*, ed. R. S. S. Fraser, pp. 361-73. Wien/New York: Springer-Verlag.
68. Linthorst, H. J. M., Meuwissen, R. L. J., Kauffmann, S., Bol, J. F. 1989. Constitutive expression of pathogenesis-related proteins PR-1, GRP and PR-5 in tobacco has no effect on virus infection. *Plant Cell* 1:285-91.
69. Lois, R., Dietrich, A., Hahlbrock, K., Schulz, W. 1989. A phenylalanine ammonia-lyase gene from parsley: structure, regulation and identification of elicitor and light-responsive cis-acting elements. *EMBO J.* 8:1641-48.
70. Lotan, T., Ori, N., Fluhr, R. 1988. *Pathogenesis Related Proteins Are Developmentally Regulated in Tobacco Flowers*. Presented at 2nd Int. Congr. Plant Mol. Biol., Jerusalem.
71. Lotan, T., Ori, N., Fluhr, R. 1989. Pathogenesis related proteins are developmentally regulated in tobacco flowers. *Plant Cell* 1:881-87.
72. Lozoya, E., Hoffmann, H., Douglas, C., Schulz, W., Scheel, D., et al. 1988. Primary structures and catalytic properties of isozymes encoded by the two 4-coumarate:CoA ligase genes in parsley. *Eur. J. Biochem.* 176:661-67.
73. Lucas, J., Henricquez, A. C., Lottspeich, F., Henschen, A., Sanger, H. L. 1985. Amino acid sequence of the "pathogenesis-related" leaf protein p14 from virus-infected tomato reveals a new type of structurally unfamiliar proteins. *EMBO J.* 4:2745-49.

- in *Plants*, ed. R. K. S. Wood, pp. 247-73. New York: Plenum.
118. Van Loon, L. C. 1985. Pathogenesis-related proteins. *Plant Mol. Biol.* 4:111-16.
  119. Van Loon, L. C. 1990. Stress proteins in infected plants. In *Plant-Microbe Interactions: Molecular and Genetic Perspectives*, ed. T. Kousuge, E. W. Nester, Van Loon, L. C., Gernitsen, Y. A. M., Ritter, C. E. 1987. Identification, purification and characterization of pathogenesis-related proteins from virus-infected Samsun NN tobacco leaves. *Plant Mol. Biol.* 9:593-609.
  121. Van Loon, L. C., Van Kammen, A. 1970. Polyacrylamide disc electrophoresis of the soluble leaf proteins from *Nicotiana glauca* var "Samsun" and "Samsun NN". II. Changes in protein constitution after infection with tobacco mosaic virus. *Virology* 40:199-211.
  122. Van Tunen, A. J., Hartman, S. A., Mur, A., Mol, J. N. M. 1989. Regulation of chalcone flavanone isomerase (CHI) gene expression in *Pistia hybridula*: the use of alternative promoters in corolla, anthers and pollen. *Plant Mol. Biol.* 12:539-51.
  123. Vera, P., Conejero, V. 1988. Pathogenesis-related proteins of tomato, P-69 as an alkaline endoprotease. *Plant Physiol.* 87:58-63.
  124. Vera, P., Hernández-Yago, J., Conejero, V. 1989. "Pathogenesis-related" (Prip14) protein. Vacuolar and apoplastic localization in leaf tissue from tomato plants infected with citrus exocortis virus; *in vitro* synthesis and processing. *J. Gen. Virol.* 70:1933-42.
  125. Verner, K., Schatz, G. 1988. Protein translocation across membranes. *Science* 241:1307-13.
  126. Voelker, T. A., Herman, E. M., Christopeels, M. J. 1989. *In vitro* mutated phytohemagglutinin genes expressed in tobacco seeds: role of glycans in protein targeting and stability. *Plant Cell* 1:95-104.
  127. Vogel, U., Meins, F., Boller, T. 1988. Co-ordinated regulation of chitinase and  $\beta$ -1,3-glucanase in bean leaves. *Planta* 174:364-72.
  128. Walter, M. H., Grima-Pettenati, J., Grand, C., Doudet, A. M., Lamb, C. J. 1988. Cinnamyl-alcohol dehydrogenase, a molecular marker specific for lignin synthesis: cDNA cloning and mRNA induction by fungal elicitor. *Proc. Natl. Acad. Sci. USA* 85:5546-50.
  129. Watson, M. E. E. 1984. Compilation of published signal sequences. *Nucleic Acids Res.* 12:5145-64.
  130. White, R. F., Rybicki, E. P., Von Wechmar, M. B., Dekker, J. L., Antoniw, J. F. 1987. Detection of PR-1 type proteins in *Amaranthaceae*, *Chenopodiaceae*, *Gramineae* and *Solanaceae* by immunoelectroblotting. *J. Gen. Virol.* 68:2043-48.
  131. Wingender, R., Rohrig, H., Hörnicke, C., Wing, D., Schell, J. 1989. Differential regulation of soybean chalcone synthase genes in plant defence, symbiosis and upon environmental stimuli. *Mol. Gen. Genet.* 218:315-22.

# LONG-DISTANCE DISPERSION OF RUST PATHOGENS

S. Nagarajan

Indian Council of Agricultural Research, Krishi Bhavan, New Delhi 110 001, India

D. V. Singh

Division of Mycology and Plant Pathology, Indian Agricultural Research Institute, New Delhi 110 012, India

KEY WORDS: trajectory, remote sensing, climatic factors, *Puccinia* path.

## INTRODUCTION

Migration and dispersal are the two main evolutionary strategies adopted by living organisms to colonize new sites. The migration of birds and mammals in search of new sites and sources of food occurs in an orderly, coordinated way with minimum wastage of progenies. Plant pathogens, on the other hand, produce enormous numbers of spores that are passively transported, scattered in all directions, and finally land on nontarget sites in uncongenial environments as well as on congenial hosts. Although the dispersion of rust urediospores may seem a wasteful exercise, this random profligacy is compensated for by the enormous numbers of spores produced per unit area. The purpose of this chapter is to describe some of the evidence and routes by which long-distance dispersal (LDD) of plant pathogens is known to occur in various parts of the world. The chapter also contains a description of new tools by which LDD can be studied and made into a predictive tool for advisory services.

## DEFINING THE TERMS

*Source area* is the suspected or proven place or warehouse from which inoculum is dispersed over both short and long distances. *Target* is a dense

# Stress Protection of Transgenic Tobacco by Production of the Osmolyte Mannitol

Mitchell C. Tarczynski,\* Richard G. Jensen, Hans J. Bohnert

The accumulation of sugar alcohols and other low molecular weight metabolites such as proline and glycine-betaine is a widespread response that may protect against environmental stress that occurs in a diverse range of organisms. Transgenic tobacco plants that synthesize and accumulate the sugar alcohol mannitol were engineered by introduction of a bacterial gene that encodes mannitol 1-phosphate dehydrogenase. Growth of plants from control and mannitol-containing lines in the absence and presence of added sodium chloride was analyzed. Plants containing mannitol had an increased ability to tolerate high salinity.

Drought, low temperature, and high salinity are environmental factors that may dramatically limit plant growth and crop productivity (1). In response to these abiotic stresses, which all disturb the intracellular water balance, many plants and bacteria synthesize and accumulate osmotically active, low molecular weight compounds such as sugar alcohols, proline, and glycine-betaine (2, 3). Collectively, these compounds have been referred to as osmolytes, osmoprotectants, or compatible solutes. Although their exact function in plants is unknown, numerous studies [(2) and references therein] suggest these osmolytes may protect the plant from abiotic stress. For example, osmolytes accumulate in plant cells in response to water or salinity stress and are subsequently degraded or lost after stress relief (2). Other studies indicate a macromolecular protective effect: in vitro incubation of osmolytes with protein extracts from plants often alleviates the adverse effects of electrolytes and temperature stress on enzymatic activity (4-7). However, several "nonprotective" roles for many of these compounds have also been suggested (8-11), such as storage products during stress. Additionally, accumulation of these compounds may be the result of a stress-induced metabolic impairment (12).

An example of such an osmolyte is the sugar alcohol mannitol, which occurs widely in plants and animals (2, 3). We previously introduced a metabolic branch point into transgenic tobacco by transformation with a 35S mtlD gene construction (13). Mannitol accumulation in leaves and roots of these transformed plants was estimated at a maximum concentration of 100 mM (13), if a cytosolic location is assumed. Here we determine whether the tobacco plant, which does not normally produce and accumulate mannitol, is protected from salin-

ity stress by this introduced osmolyte.

The growth of 6-week-old plants in the absence and presence of added NaCl was assessed. We evaluated plant growth by measuring the percent of change in height and in fresh weight (14). Initial heights and weights for all plants of each experiment were measured just before NaCl addition (15). For nonstressed plants, final measurements were taken at the onset of flowering (after 9 to 10 days of culture). Final heights and fresh weights for stressed plants were measured after 14 and 30 days of culture because these plants did not flower. No differences were observed between control and mannitol-containing plants after 10 days of NaCl treatment. Plants were evaluated from five different lines: the two controls, SR1, wild-type tobacco cv. SR1 and SR1<sup>km</sup>, kanamycin-resistant tobacco, which was transformed with a gene construction identical to that used for mannitol expression (13) except for the absence of the mtlD gene, and three lines of mannitol-containing transformants, 1-mtl, 2-mtl,

and 3-mtl (16). Plants were hydroponically cultured under controlled conditions (17). All plants from mtl lines contained mannitol, whereas mannitol was not detected in plants from control lines (13).

When cultured without added NaCl, both the control and mtl plants increased about six- to sevenfold in height and more than twofold in fresh weight over the 9- to 10-day interval before flowers appeared (Table 1). Statistical analyses of percent change in height and fresh weight [single degree of freedom contrasts from a combined analysis of variance (ANOVA) of two experiments] indicated no significant differences in these variables between the two control lines or between the control lines and the three mtl lines ( $P = 0.17$  to  $0.63$ ). Thus, neither the transformation protocol nor mannitol accumulation influenced plant growth in the absence of excess NaCl.

To determine the NaCl concentrations at which mannitol may affect growth, we conducted pilot studies with 6-week-old plants exposed to various concentrations of NaCl (100 to 300 mM) for 30 days. Primarily on the basis of visual analysis, growth of plants containing mannitol could be distin-

**Table 1.** Growth of tobacco plants in the absence of added NaCl. All data from two separate experiments are shown. Seeds were germinated and plants were cultured in an equal mixture of vermiculite, potting soil, and sand under greenhouse conditions for 3 weeks. Intact plants were then transferred to a hydroponics system and cultured in a growth room for ~3 weeks ( $\pm 1$  day) (17). Height of the aerial portion (in centimeters) and total fresh weight of individual plants (in grams) were measured (initial measurement, *I*). These two traits were measured again for the same plants just before flowering, 9 to 10 days later (final measurement, *F*). Roots were blotted dry before weight measurements were taken. Percent change was calculated as  $[(F/I) - 1] \times 100$ . Lines evaluated (16): SR1, wild type; SR1<sup>km</sup>, kanamycin-resistant; and 1-mtl, 2-mtl, and 3-mtl (three individual transformants), kanamycin-resistant and expressing 35S mtlD (13). Evaluations were performed blindly. Data are mean values  $\pm$  SEM; *n*, total number of plants.

Plant line	Height (% change)	Fresh weight (% change)	<i>n</i>
<i>Experiment 1</i>			
SR1	748.8 $\pm$ 110.9	301.5 $\pm$ 29.4	8
SR1 <sup>km</sup>	718.1 $\pm$ 95.6	248.0 $\pm$ 20.1	6
1-mtl	791.4 $\pm$ 78.6	330.8 $\pm$ 26.3	8
2-mtl	871.9 $\pm$ 91.4	276.1 $\pm$ 31.6	6
3-mtl	615.7 $\pm$ 36.2	270.6 $\pm$ 17.6	8
<i>Experiment 2</i>			
SR1	577.3 $\pm$ 26.1	199.7 $\pm$ 5.7	8
SR1 <sup>km</sup>	546.8 $\pm$ 33.2	202.7 $\pm$ 6.5	8
1-mtl	560.8 $\pm$ 48.1	200.5 $\pm$ 13.1	8
2-mtl	643.9 $\pm$ 29.5	250.3 $\pm$ 9.8	8
3-mtl	582.3 $\pm$ 40.3	203.0 $\pm$ 12.4	8



**Fig. 1.** Control and mannitol-containing tobacco plants after 30 days of culture in the presence of added NaCl (250 mM). Plants were cultured as described (17). Left plant, wild type (SR1); right plant, 35S mtlD transformant (1-mtl).

Department of Biochemistry, University of Arizona, Tucson, AZ 85721.

\*To whom correspondence should be addressed at 5 Prime  $\rightarrow$  3 Prime, Inc., 5603 Arapahoe, Boulder, CO 80303.

guished from that of control plants at high salinity (for example, 250 mM). Thus, we pursued this initial observation by assessing the growth of plants cultured under the extreme, or shock, condition of 250 mM NaCl.

After 14 days of treatment with 250 mM NaCl, plant growth from both control and mtl lines was severely inhibited (Table 2). Stem elongation of salt-treated plants was generally less than 10% of that for the

nonstressed plants after 9 to 10 days of culture. Additionally, plants typically did not flower after 14 days of stress. The fresh weight of plants from most lines decreased because of necrosis and dehydration of lower leaves. A comparison of the different lines for each experiment shows that after 14 days of stress there were no clear differences between the growth of plants from control and mtl lines.

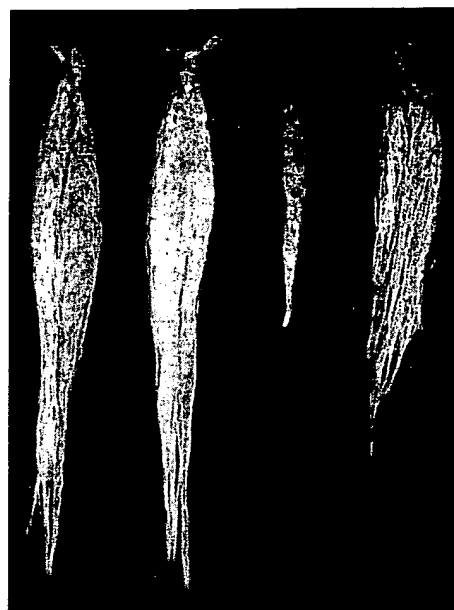
After 30 days of exposure to high salinity, however, this difference could be distinguished, with plants that contained mannitol showing increased tolerance (Table 2). In both experiments, plants from mtl lines had decreased weight loss relative to that for plants of control lines. Additionally, mtl transformants increased in height a mean of 80% (range of 53 to 121%) compared with their height at 14 days. Conversely, control plants increased in height only a mean of 22% (range of 3 to 32%) over the same interval. Statistical analyses (single degree of freedom contrasts from a combined ANOVA for both experiments) documented significant differences between control and mtl lines in mean height ( $P = 0.0006$ ) and fresh weight ( $P = 0.0001$ ). No such differences were observed between the two control lines. Thus, on the basis of height and fresh weight, plants that contained mannitol showed increased tolerance to high salinity relative to control plants. Variability in results (for example, with 3-mtl plants) may reflect subtle differences in plant development among the lines just before NaCl treatment or in the experimental conditions (18).

Control and transgenic plants also appeared different after 30 days of stress with 250 mM NaCl (Figs. 1 and 2 and Table 3). With all plants, roots exposed to 250 mM NaCl were apparently unable to survive

treatment after about 10 days and appeared dark brown in color (salt-stressed control roots, Fig. 2). However, plants that contained mannitol often produced new roots and then new leaves (Figs. 1 and 2 and Table 3). New root and leaf growth occurred in a mean of 75% (range of 60 to 88%) of plants from mtl lines, whereas such growth occurred in only a mean of 33% (range of 13 to 50%) of plants from control lines (Table 3). These new roots typically appeared after about 14 days of NaCl treatment. Flowering (Fig. 1) was observed in 73% of plants that generated new roots and leaves. Although it is not clear how intracellular mannitol accumulation may lead to new root growth, the osmolyte may affect processes at the cellular level that are involved in formation of new roots.

The increased production of new roots and the subsequent emergence of new leaves and then flowers indicate that the relative increases in height and fresh weight of plants containing mannitol are due to new growth rather than to a reallocation of resources. Additional experiments with dry weight analyses (which were not carried out in our time course studies because of their destructive nature) should confirm this observation.

Sugar alcohols may contribute to tolerance at the cellular level by adjustment of the cytosolic osmotic potential when the concentration of electrolytes is lower in the cytosol than in the vacuole (2). These compounds may also protect membranes and proteins in the presence of high concentrations of electrolytes (4, 19). An assumption in both mechanisms is that the osmolytes are localized in the cytosol. Our data do not distinguish between these mechanisms. Depending on the cell type, its normal responses to stress, and environ-



**Fig. 2.** Roots from control and mannitol-containing tobacco plants after 30 days of culture in the absence and presence of added NaCl (250 mM). Plants were cultured as described (17). From left to right: wild-type plant (SR1) cultured without added NaCl, 35S mtlD transformant (1-mtl) cultured without added NaCl, SR1 plant cultured in 250 mM NaCl, and 1-mtl plant cultured in 250 mM NaCl.

**Table 2.** Growth of tobacco plants in the presence of 250 mM NaCl. All data from two separate experiments are shown. Plant height and fresh weight were measured from ~6-week-old plants ( $\pm 1$  day) just before salt addition (added all at once) and then at 14 and 30 days after addition. Evaluations were performed blindly. Procedures and abbreviations as in Table 1. Data are mean values  $\pm$  SEM.

Plant line	Height (% change)		Fresh weight (% change)		n
	14 days	30 days	14 days	30 days	
Experiment 1					
SR1	45.9 ± 3.9	60.3 ± 9.5	-14.0 ± 5.8	-48.4 ± 11.1	8
SR1 <sup>km</sup>	66.9 ± 8.5	68.7 ± 10.8	-9.5 ± 9.7	-44.7 ± 16.8	8
1-mtl	59.4 ± 4.2	113.4 ± 29.8	-1.7 ± 12.2	-7.7 ± 24.9	6
2-mtl	54.4 ± 9.6	83.3 ± 14.9	-1.6 ± 4.9	-14.8 ± 10.4	8
3-mtl	74.5 ± 10.2	129.3 ± 12.0	+16.3 ± 5.2	+27.6 ± 12.0	8
Experiment 2					
SR1	23.1 ± 2.8	28.0 ± 3.0	-9.1 ± 5.7	-34.0 ± 8.2	10
SR1 <sup>km</sup>	36.5 ± 3.5	48.2 ± 5.4	-7.3 ± 5.6	-36.2 ± 7.5	10
1-mtl	29.4 ± 3.1	55.1 ± 7.1	-6.5 ± 3.6	-18.4 ± 10.4	10
2-mtl	19.2 ± 2.8	42.5 ± 7.3	-6.9 ± 5.6	-14.7 ± 9.3	10
3-mtl	29.0 ± 2.3	44.8 ± 3.4	-4.3 ± 3.4	-14.8 ± 9.8	10

**Table 3.** Visual assessment of new root and leaf growth in tobacco plants after 30 days of exposure to 250 mM NaCl. All data from two separate experiments are shown. Evaluations were performed blindly. Procedures and abbreviations as in Table 1.

Plant line	Plants with new roots and leaves	n
<i>Experiment 1</i>		
SR1	1	8
SR1 <sup>km</sup>	2	8
1-mtl	4	6
2-mtl	6	8
3-mtl	7	8
<i>Experiment 2</i>		
SR1	5	10
SR1 <sup>km</sup>	4	10
1-mtl	8	10
2-mtl	8	10
3-mtl	6	10



mental conditions, a contribution in tobacco cells of 100 mM mannitol may be consistent with both mechanisms. However, uniform distribution of mannitol at the subcellular level would likely preclude cytosolic osmoregulation. Determination of both the subcellular location of mannitol and the concentrations of the sugar alcohol in tissues will help to distinguish between these two mechanisms.

Alternatively, mannitol may metabolically predispose tobacco cells to stress tolerance. Thus, the cellular accumulation of mannitol, which is normally foreign in tobacco cells, may increase the response of metabolic pathways normally involved in stress tolerance, which thereby allow cells to withstand stress.

On the basis of pilot experiments, we studied the growth response of plants of a defined age that were exposed to the extreme, or shock, condition of 250 mM NaCl. Additional experiments that vary plant age (development) and NaCl concentration (added all at once or incrementally), as well as experiments that investigate the effects of other environmental stresses such as drought and cold, will help to explain the function of mannitol in higher plants. We have demonstrated that for tobacco, the presence of mannitol in vivo protects against high salinity. Sugar alcohol accumulation may also enhance stress tolerance in other plants.

## REFERENCES AND NOTES

1. J. S. Boyer, *Science* 218, 443 (1982).
2. J. A. Hellebust, *Annu. Rev. Plant Physiol.* 27, 485 (1976).
3. P. H. Yancey, M. E. Clark, S. C. Hand, R. D. Bowler, G. N. Somero, *Science* 217, 1214 (1982).
4. A. Pollard and R. G. Wyn Jones, *Planta* 144, 291 (1979).
5. Y. Manetas, Y. Petropoulou, G. Karabourniotis, *Plant Cell Environ.* 9, 145 (1986).
6. J. P. Krall, G. E. Edwards, C. S. Andreo, *Plant Physiol.* 89, 280 (1989).
7. S. Laurie and G. R. Stewart, *J. Exp. Bot.* 41, 1415 (1990).
8. H. Greenway and R. Munns, *Annu. Rev. Plant Physiol.* 31, 149 (1980).
9. D. Aspinall and L. G. Paleg, in *The Physiology and Biochemistry of Drought Resistance in Plants*, L. G. Paleg and D. Aspinall, Eds. (Academic Press, New York, 1981), pp. 205-241.
10. R. L. Bielecki, in *Encyclopedia of Plant Physiology*, F. A. Loewus and W. Tanner, Eds. (Springer-Verlag, Berlin, 1982), new series, vol. 13A, pp. 158-192.
11. A. D. Hanson and W. D. Hitz, *Annu. Rev. Plant Physiol.* 33, 163 (1982).
12. R. G. Wyn Jones and J. Gorham, in *Encyclopedia of Plant Physiology: Responses to the Environment*, O. L. Lange, P. S. Nobel, C. B. Osmond, H. Ziegler, Eds. (Springer-Verlag, Berlin, 1983), vol. 12C, pp. 35-58.
13. M. C. Tarczynski, R. G. Jensen, H. J. Bohnert, *Proc. Natl. Acad. Sci. U.S.A.* 89, 2600 (1992).
14. Data for percent changes in height and fresh weight collected over time for the same individual plants were used to determine the inherent variability in initial measurements and were used in all statistical analyses. Data for percent changes in both height and fresh weight were shown to be

normally distributed ( $P \leq 0.01$ ) by calculation of the Shapiro-Wilk statistic that permitted our evaluation with ANOVA.

15. Initial heights/weights for experiment 1; and those from experiment 2, respectively, in percent change, for each line. Plants not receiving NaCl: SR1,  $6.1 \pm 0.4/21.6 \pm 1.1$ ,  $7.3 \pm 0.5/26.9 \pm 0.9$ ; SR1<sup>km</sup>,  $6.7 \pm 0.5/27.2 \pm 1.2$ ,  $7.6 \pm 0.3/27.6 \pm 0.9$ ; 1-mtl,  $6.4 \pm 0.6/23.7 \pm 1.5$ ,  $7.3 \pm 0.6/26.0 \pm 1.1$ ; 2-mtl,  $5.9 \pm 0.3/27.3 \pm 1.1$ ,  $6.4 \pm 0.4/24.9 \pm 0.8$ ; 3-mtl,  $6.3 \pm 0.3/21.9 \pm 1.9$ ,  $7.1 \pm 0.5/27.1 \pm 1.5$ . Plants subsequently receiving NaCl treatment: SR1,  $7.1 \pm 0.4/27.9 \pm 2.0$ ,  $7.4 \pm 0.2/26.1 \pm 1.3$ ; SR1<sup>km</sup>,  $9.4 \pm 0.4/30.8 \pm 1.1$ ,  $7.1 \pm 0.3/25.5 \pm 1.0$ ; 1-mtl,  $7.4 \pm 0.4/26.4 \pm 1.4$ ,  $7.2 \pm 0.3/28.1 \pm 0.7$ ; 2-mtl,  $6.8 \pm 0.3/27.5 \pm 1.4$ ,  $6.6 \pm 0.3/27.4 \pm 0.7$ ; 3-mtl,  $7.3 \pm 0.4/26.3 \pm 1.6$ ,  $7.1 \pm 0.2/27.9 \pm 0.8$ . Values are mean  $\pm$  SEM; for number of plants evaluated; see Tables 1 and 2.
16. For lines that express 35S mtlD, 30 primary transformants were chosen randomly and self-pollinated, and the progeny were selected for single-locus inserts on the basis of expression of kanamycin resistance. Of the lines that contain single locus inserts, three were randomly chosen (1-mtl, 2-mtl, 3-mtl), self-pollinated again, and determined to be homozygous on the basis of expression of kanamycin resistance. This procedure was also used for control SR1<sup>km</sup>, except that five primary transformants were chosen. Lines analyzed were SR1, wild type; SR1<sup>km</sup>, kanamycin-resistant; and 1-mtl, 2-mtl, and 3-mtl, three different kanamycin-resistant lines that express 35S mtlD.
17. The hydroponics system consisted of two genetically identical plants cultured in containers with half-strength Hoagland's nutrient solution (5 li-

ters). (For Figs. 1 and 2, plants from SR1 and 1-mtl were cultured together in 250 mM NaCl in hydroponic containers.) The nutrient solution was changed every 2 weeks. Plants were inserted into the nutrient solution through two holes in the container lid. The growth room housed eight to ten plots of five randomly placed containers (one for each line). For each experiment, half of the plots were cultured in the absence of added NaCl and the other half were cultured in the presence of 250 mM NaCl. To minimize variability in growth and development of plants before stress, all plants (for the data reported in Tables 1 to 3) were approximately the same age (6 weeks) and were selected from larger groups for inclusion in each experiment for their similar heights and weights. For the two experiments, plants were visually identical before NaCl treatment. To eliminate potential experimenter bias, evaluations were performed blindly. Growth-room conditions were 18° and 25°C, night and day temperatures, respectively; photoperiod, 12 hours light (500  $\mu$ E m<sup>-2</sup> s<sup>-1</sup>).

18. R. L. Mott and F. C. Steward, *Ann. Bot.* 36, 621 (1972); *ibid.*, p. 641; *ibid.*, p. 655; *ibid.*, p. 915.
19. B. Schobert, *J. Theor. Biol.* 68, 17 (1977).
20. We thank D. M. Vernon, J. C. Thomas, J. W. O'Leary, B. Larkins, and F. Lehle for manuscript comments; P. Adams for high-performance liquid chromatography; S. Smith for statistical analyses and manuscript comments; J. Orr and J. Class for technical assistance; and I. H. Tarczynski for patience and technical support. Supported by U.S. Department of Agriculture grant CSRS 91-371006539 (M.C.T. and R.G.J.) and Department of Energy Grant DE-FG02-92ER20066.

8 July 1992; accepted 6 November 1992

## Altered Specificity of DNA-Binding Proteins with Transition Metal Dimerization Domains

Bernard Cuenoud and Alanna Schepartz\*

The bZIP motif is characterized by a leucine zipper domain that mediates dimerization and a basic domain that contacts DNA. A series of transition metal dimerization domains were used to alter systematically the relative orientation of basic domain peptides. Both the affinity and the specificity of the peptide-DNA interaction depend on domain orientation. These results indicate that the precise configuration linking the domains is important; dimerization is not always sufficient for DNA binding. This approach to studying the effect of orientation on protein function complements mutagenesis and could be used in many systems.

Active sites of proteins are typically composed of recognition elements guided into proximity and appropriate orientation by the native protein fold. Individual recognition elements may be remote in primary structure or may be located on different polypeptide chains in multisubunit proteins. With site-directed mutagenesis, amino acids that constitute individual recognition elements can be changed without affecting the overall orientation of the recognition domain. Yet present technology does not allow predictable and routine changes in the orientation of the domains themselves.

Department of Chemistry, Yale University, New Haven, CT 06511-8118.

\*To whom correspondence should be addressed.

The transcriptional activator protein GCN4 is one of a large family of DNA-binding proteins identified by a bZIP structural motif (1); this motif contains a DNA contact domain characterized by conserved basic and hydrophobic residues (b domain), and a dimerization domain identified by a heptad repeat of leucine residues (ZIP domain) (2, 3). The two domains are separated by a six-amino acid linker whose length, but not sequence is conserved across bZIP families (1). Previous work has demonstrated that the active DNA-binding entity is generated when the ZIP domains of two protein monomers assemble (4) into a parallel coiled coil (5, 6). The scissors grip (1) and induced helical fork (7) models propose that the coiled coil, the natural dimeriza-



WGA  


# Molecular Cloning and Physical Mapping of the *otsBA* Genes, Which Encode the Osmoregulatory Trehalose Pathway of *Escherichia coli*: Evidence that Transcription Is Activated by KatF (AppR)

INGA KAASEN, PÅL FALKENBERG, OLAF B. STYRVOLD, AND ARNE R. STRØM\*  
*The Norwegian College of Fishery Science, University of Tromsø, N-9000 Tromsø, Norway*

Received 16 July 1991/Accepted 26 November 1991

It has been shown previously that the *otsA* and *otsB* mutations block osmoregulatory trehalose synthesis in *Escherichia coli*. We report that the transcription of these osmoregulated *ots* genes is dependent on KatF (AppR), a putative sigma factor for certain stationary phase- and starvation-induced genes. The transcription of the osmoregulated *bet* and *proU* genes was not *katF* dependent. Our genetic analysis showed that *katF* carries an amber mutation in *E. coli* K-12 and many of its derivatives but that *katF* has reverted to an active form in the much-used strain MC4100. This amber mutation in *katF* leads to strain variations in trehalose synthesis and other *katF*-dependent functions of *E. coli*. We have performed a molecular cloning of the *otsBA* genes, and we present evidence that they constitute an operon encoding trehalose-6-phosphate phosphatase and trehalose-6-phosphate synthase. A cloning and restriction site analysis, performed by comparing the cloned fragments with the known physical map of the *E. coli* chromosome, revealed that the *otsBA* genes are situated on a 2.9-kb *Hind*III fragment located 8 to 11 kb clockwise of *tar* (41.6 min).

Trehalose, a nonreducing disaccharide of glucose, is a stress metabolite in various organisms (29). *Saccharomyces cerevisiae* accumulates trehalose when exposed to an elevated temperature of growth (3, 24) or to hazardous chemical agents such as ethanol, copper sulfate, or hydrogen peroxide (3). Rhizobia accumulate trehalose when stressed with low-oxygen (e.g., 1%) tension, regardless of the composition of the growth medium (23). Many phototrophic and heterotrophic bacteria, including *Escherichia coli*, accumulate trehalose in response to osmotic stress (18, 31, 45, 50). Trehalose is shown to preserve the function and integrity of biological membranes exposed to conditions of low water activity (14) and to confer desiccation tolerance to yeasts (24), to spores of *Streptomyces* sp. (36), and to nematodes (14); frost tolerance to insects (2) and yeasts (22); and osmotic tolerance to *E. coli* (19). In yeasts, trehalose accumulation during growth in liquid culture coincides with an increased plating efficiency on agar plates of low water activity (34).

In *E. coli*, the osmoregulatory trehalose pathway consists of a trehalose-6-phosphate synthase which converts UDP-glucose and glucose-6-phosphate to trehalose-6-phosphate and a phosphatase which dephosphorylates this metabolic intermediate (reference 19 and this study). Two insertion mutations, named *otsA* and *otsB*, which block the synthesis of the synthase, have previously been mapped to 42 min, but the trehalose-6-phosphate phosphatase activity of these mutants was not reported (19). However, a point mutation named *otsP* which causes accumulation of trehalose-6-phosphate in stressed cells, presumably because of a defective phosphatase, was mapped near *otsA* (27). This mutation appears to be allelic with *otsB* (reference 27 and this study).

Trehalose accumulation in stressed cells of *E. coli* is regulated at several levels. Experiments with *lac* fusions have shown that the *otsA* and *otsB* genes are transcription-

ally activated by osmotic stress, and biochemical data have shown that the synthase is activated by potassium glutamate (19). At the cellular level, trehalose accumulation is regulated by a futile cycle. Trehalose is overproduced in the cytoplasm, and the excess is excreted and split to glucose by a periplasmic trehalase (TreA) and then reutilized (52). Similarly to the *ots* genes, the *treA* gene is transcriptionally activated by osmotic stress but not by external trehalose (8). *treA* maps at 26 min (8). Rod et al. (47) reported that *E. coli* K-12 and most of its derivatives carry an amber mutation that leads to decreased accumulation of trehalose in osmotically stressed cells. Styrvold and Strøm (52) showed that this amber mutation was not in the *otsA* and *otsB* genes themselves but that it caused decreased transcription of these genes.

Trehalose can serve as the sole source of energy and carbon in *E. coli*. At high osmolarity, trehalose utilization depends on the TreA activity (20). But at low osmolarity, trehalose induces a transport protein, EII<sup>Tre</sup> (TreB), and a pathway consisting of at least a catabolic trehalose-6-phosphate phosphatase (TreE) and an amylorehalase (TreC). *treB*, *treC*, and presumably *treE* map at 96.5 min (9). Their internal inducer is trehalose-6-phosphate; therefore, these genes are not expressed under osmotic stress when the osmoregulatory trehalose-6-phosphate phosphatase is operative (27). A mutation which seems to influence expression of *otsA*, *otsP* (*otsB*), and *treC*, but not *treB*, has been mapped to 84 min, but the nature of this gene remains unclear (27).

In this investigation, we have cloned the *otsBA* genes. We report that they constitute an operon which encodes the phosphatase and the synthase, respectively, of the osmoregulatory trehalose pathway. The amber mutation which influences *otsBA* expression was mapped at 59 min and shown to be in *katF*, which encodes a putative sigma factor (39) for starvation- and stationary phase-induced genes (30, 37).

\* Corresponding author.

## MATERIALS AND METHODS

**Bacterial strains and growth conditions.** The bacterial strains used are listed in Table 1. The growth media (medium 63 [38], low-osmolarity minimal medium [LOM] [11], and LB medium [38]) and the growth conditions used were exactly as described previously by us (52).

**Genetic procedures and strain constructions.** Conjugation by mating on agar plates and P1 transduction were performed as described by Miller (38). When transferring Tn10 insertions and *lacZ* fusions generated with  $\lambda$  *placMu*, we selected for resistance against tetracycline and kanamycin, respectively. Tn10 was deleted by using the positive selection method described by Bochner et al. (7). The presence of a *recA* mutation was tested by determining UV sensitivity as described previously (35). For strains carrying *otsA-lacZ* or *otsB-lacZ* operon fusions, the *otsX* (*katF*) genotype was routinely checked by scoring their Lac phenotype by growth on medium 63 with lactose as an energy source (see below). The pedigrees of the strains constructed in this study are given in Table 1. Further description of strain construction is given in this section and in Results.

A culture of MC4100 carrying a random selection of Tn10 insertions was prepared by infecting MC4100 with the phage  $\lambda$  NK561 essentially as described previously (49). After the transposition, the cells were spread on 50 agar plates containing LB medium with 25  $\mu$ g of tetracycline per ml and 4 mM sodium citrate. Approximately 50,000 Tc<sup>r</sup> colonies representing individual transposition events were pooled and used for production of P1 lysate.

We verified that our strain of MC4100 did not carry an amber suppressor in the following way. A Tn10 insertion was introduced near an amber mutation by transducing BL50 [*metF*(Am)] with a lysate prepared from GD1 (*zih::Tn10*) by selecting for Tc<sup>r</sup> and screening for Met<sup>-</sup>. A P1 lysate of a resulting strain, PFF1 [*metF*(Am) *zih::Tn10*], was then used to transduce MC4100 to Tc<sup>r</sup>. Of 50 Tc<sup>r</sup> transductants tested, 19 were Met<sup>-</sup>, showing that the *metF*(Am) mutation was not suppressed in MC4100.

FF1005 was constructed by conjugational crossing of JC10240 with MC4100 by selecting for *srl::Tn10* (Tc<sup>r</sup>) and screening for *recA* (UV<sup>s</sup>). By also checking that the transductant gave an osmotically tolerant phenotype on agar plates with medium 63 with 0.5 M NaCl added, we ensured that FF1005 carried the *otsX*<sup>MC4100</sup> allele of the recipient and not the *otsX* allele of the donor.

In order to verify that LCB107 carries a *supE* mutation and to construct a suppressor-free LCB107 derivative, a P1 lysate prepared from BL50 [*metB*<sup>+</sup> *metF*(Am)] was first used to transduce the LCB107-derived IK5 (*metB1 metF<sup>+</sup> supE?*) to Met<sup>+</sup>. As *metB* and *metF* are neighbor genes on the chromosome (4), Met<sup>+</sup> transductants that carry an amber suppressor could have either a *metB*<sup>+</sup> *metF*(Am) or a *metB*<sup>+</sup> *metF*<sup>+</sup> genotype. To investigate their genotype, four transductants from this cross were transduced to Tc<sup>r</sup> with a P1 lysate prepared from JW1071 (*sup<sup>0</sup> zbf::Tn10*), which carries a Tn10 marker within cotransductional distance of *glnV*, i.e., the native *supE* gene (4). The progeny from each of the four crosses displayed both Met<sup>+</sup> (30%) and Met<sup>-</sup> (70%) genotypes. This showed that their parental strains were *supE metF*(Am). One of the Met<sup>+</sup> transductants of strain IK5 was named IK58. Strain IK59 was a *sup<sup>0</sup> metF*(Am) derivative of IK58.

To introduce the *otsX*<sup>MC4100</sup> allele into an LCB107 background, we first transduced the LCB107 derivative IK58 with a P1 lysate prepared from IK68 (*cysC mutS::*

Tn10), selecting for Tc<sup>r</sup> and screening for Cys<sup>-</sup>. The resulting strain, IK60, was transduced with a P1 lysate prepared from FF1112 (*cysC<sup>+</sup> mutS<sup>+</sup> otsX<sup>+</sup> MC4100), selecting for Cys<sup>+</sup> and screening for Tc<sup>r</sup>. As *otsX* is located between the *cysC* and the *mutS* genes (see below), a Cys<sup>+</sup> Tc<sup>r</sup> transductant such as IK61 should carry the *otsX*<sup>MC4100</sup> allele.*

**General recombinant DNA procedures and cloning of *otsA* and *otsB*.** Transformations were done by the method of Chung et al. (13) or a standard CaCl<sub>2</sub> method. Plasmid isolation, restriction cleavage of DNA, and ligation were done by standard methods.

Restriction maps of the chromosomally derived parts of the constructed plasmids are shown in Fig. 2, and the plasmids are listed in Table 1. A gene library of strain CSH7 was constructed in the pBR322-derived cosmid vector cos4 (46) as described by Andresen et al. (1). Plasmids were introduced into the osmotically sensitive strain FF4169 (*otsA::Tn10*), and osmotically tolerant, trehalose-producing clones were selected by plating on medium 63 with 0.45 M NaCl added. The plasmid pFF1, selected in this way, contained a 43-kb chromosomal fragment (see Fig. 2). The left end of the fragment is proximal to the *SalI* site of the vector. The plasmid pFF1 was maximally shortened by *SalI* to give pFF101 and by *Bam*HI to give pFF102. A 2.9-kb *Hind*III fragment from pFF101 was subcloned into the *Hind*III site of pACYC184 (12) to give the subclone pFF106. The left *Hind*III site of the chromosomal insert of pFF106 (see Fig. 2) is proximal to the *Clal* site of the vector. *Hind*III shortening of pFF102 yielded pFF109. In the construction of pFF114, plasmid pFF106 was shortened by *Eco*RV so that 0.8 kb of the insert and 0.2 kb of the vector were deleted.

**Trehalose-6-phosphate synthase and trehalose-6-phosphate phosphatase activities.** The trehalose-6-phosphate synthase and phosphatase enzyme activities were measured in vitro by using cells permeabilized with toluene (19). In the synthase assay, trehalose-6-phosphate formed from UDP-glucose and glucose-6-phosphate was dephosphorylated by a treatment with alkaline phosphatase (19, 52) and the trehalose formed was trimethylsilylated and determined by gas chromatography exactly as described previously (19). The standard reaction mixture for determination of the phosphatase activity contained (in a 150- $\mu$ l volume) the following: 1.5  $\mu$ mol of trehalose-6-phosphate, 5  $\mu$ mol of Tris-hydrochloride (pH 7.4), 0.4  $\mu$ mol of MgCl<sub>2</sub>, and 125 to 500  $\mu$ g of cell protein. The reaction mixture was incubated up to 12 min at 37°C, and the reaction was terminated by heating for 5 min in a boiling-water bath. Sucrose (0.25  $\mu$ mol in a 25- $\mu$ l volume) was then added as an internal standard, and denatured protein was removed by centrifugation. For desalting, a sample of 150  $\mu$ l was applied to a column (0.5 by 2 cm) packed with equal amounts of Dowex 50X4-200 in H<sup>+</sup> form and Dowex 1X8-400 in formate form. Free sugars were washed through the column with 1 ml of water, and the eluate was freeze-dried. Gas chromatographic determination of trimethylsilylated trehalose was then performed exactly as described previously (19). One unit of synthase or phosphatase activity equals 1 nmol of trehalose produced per min at 37°C.

**Other methods and special chemicals.** Cell protein was determined by the biuret method as modified for whole bacterial cells (21). The  $\beta$ -galactosidase activity of *lacZ* fusion mutants was determined quantitatively by using cells treated with chloroform and sodium dodecyl sulfate as described previously (19, 38). One unit of  $\beta$ -galactosidase activity equals 1 nmol of *o*-nitrophenol produced per min at 28°C. Catalase activity and glycogen accumulation were

TABLE 1. *E. coli* strains, bacteriophages, and plasmids

Strain, bacteriophage, or plasmid	Description <sup>a</sup>	Construction or source <sup>a</sup>
<b>Strains</b>		
AT2427	Hfr <i>cysJ43 relA1 thi-1 spoT1</i> $\lambda^-$	CGSC 4502
BL50	F <sup>-</sup> <i>thr-1(Am) leuB6 his-4 metF159(Am) eda-50 rpsL136 thi-1 uru-14 mtl-1 xyl-5 tsx-78 lacY</i>	B. Lowsky
CSH7	F <sup>-</sup> <i>lacY(Am) rpsL thi</i>	38
DC906	F <sup>-</sup> <i>relA1 spoT1 metB1 fuc::Tn10 cys153(Am) otsX</i>	D. P. Clark (47)
DC931	W1485 <i>cys153(Am) lacZ125(Am)</i>	D. P. Clark (47)
DC943	W1485 <i>cys153(Am) lacZ125(Am) supE42</i>	D. P. Clark (47)
ES1481	F <sup>-</sup> <i>lacZ53(Am) mutS215::Tn10 thrA36 rha-5 metB1 deoC2 [NtrnD-rnnE]1? <math>\lambda^-</math> otsX<sup>-</sup></i>	CGSC 7049
FF1005	MC4100 <i>recA56 srl-300::Tn10 otsX<sup>-</sup></i> MC4100	MC4100 $\times$ JC10240
FF1112	MC4100 $\Phi$ otsB-lacZ18 <i>otsX<sup>-</sup></i> MC4100	19
FF1608	MC4100 $\Phi$ otsB-lacZ19 <i>otsX<sup>-</sup></i> MC4100	19
FF2032	MC4100 $\Phi$ otsA-lacZ17 <i>otsX<sup>-</sup></i> MC4100	19
FF4012	MC4100 <i>otsA1::Tn10</i> $\Phi$ otsB-lacZ19 <i>otsX<sup>-</sup></i> MC4100	P1(FF1608) $\times$ FF4169
FF4019	MC4100 <i>ire-1 <math>\Delta</math>zcf-229::Tn10</i> <i>otsX<sup>-</sup></i> MC4100	19
FF4025	MC4100 <i>ire-1 <math>\Delta</math>zcf-229::Tn10</i> $\Phi$ otsB-lacZ19 <i>otsX<sup>-</sup></i> MC4100	19
FF4026	MC4100 <i>ire-1 <math>\Delta</math>zcf-229::Tn10</i> $\Phi$ otsA-lacZ17 <i>otsX<sup>-</sup></i> MC4100	P1(FF2032) $\times$ FF4019
FF4031	MC4100 $\Delta$ otsA1::Tn10 $\Phi$ otsB-lacZ18 <i>otsX<sup>-</sup></i> MC4100	19
FF4035 <sup>c</sup>	MC4100 $\Delta$ otsA1 $\Phi$ otsB-lacZ1 <i>otsX<sup>-</sup></i> MC4100	Tc <sup>r</sup> Km <sup>r</sup> of FF4012
FF4037	MC4100 $\Delta$ otsA1::Tn10 $\Phi$ otsB-lacZ18 <i>otsX<sup>-</sup></i> MC4100	P1(UE5) $\times$ FF4031
FF4049	MC4100 $\Delta$ otsA1::Tn10 $\Phi$ otsB-lacZ18 <i>otsX<sup>-</sup></i> MC4100	Tc <sup>r</sup> of FF4037
FF4050	MC4100 $\Delta$ otsA1::Tn10	P1(FF1005) $\times$ FF4049
FF4055 <sup>c</sup>	MC4100 $\Delta$ otsA1 $\Phi$ otsB-lacZ1 <i>cys-95::Tn10 otsX<sup>-</sup></i> W1485	P1(1K26) $\times$ FF4035
FF4056 <sup>c,d</sup>	MC4100 $\Delta$ otsA1 $\Phi$ otsB-lacZ1 $\Delta$ cys-95::Tn10 <i>otsX<sup>-</sup></i> W1485?	Tc <sup>r</sup> of FF4055
FF4057 <sup>c,d</sup>	MC4100 $\Delta$ otsA1 $\Phi$ otsB-lacZ1 $\Delta$ cys-95::Tn10 <i>otsX<sup>-</sup></i> W1485?	P1(FF4171) $\times$ FF4056
FF4169	MC4100 <i>otsA1::Tn10</i>	19
FF4171	MC4100 <i>ireA::Tn10</i>	52
GD1	MC4100 <i>zih-730::Tn10 glpR</i>	B. Lowsky
IK1	MC4100 $\Phi$ otsB-lacZ18 <i>fuc::Tn10 otsX<sup>-</sup></i> MC4100	P1(DC906) $\times$ FF1112
IK2	W1485 $\Delta$ lac <i>otsX<sup>-</sup></i> W1485	Tc <sup>r</sup> of OS18
IK3	W1485 $\Delta$ lac $\Phi$ otsA-lacZ17 <i>otsX<sup>-</sup></i> W1485	P1(FF2032) $\times$ IK2
IK4	W1485 $\Delta$ lac $\Phi$ otsB-lacZ18 <i>otsX<sup>-</sup></i> W1485	P1(FF1112) $\times$ IK2
IK5	LCB107 $\Delta$ lac $\Phi$ otsA-lacZ17 <i>otsX<sup>-</sup></i> LCB107 <i>supE44</i>	Tc <sup>r</sup> of OS36
IK6	MC4100 $\Phi$ otsB-lacZ18 <i>srl-300::Tn10 otsX<sup>-</sup></i> MC4100	P1(1K10) $\times$ FF1112
IK10	W1485 $\Delta$ lac $\Phi$ otsB-lacZ18 <i>srl-300::Tn10 cys153(Am) otsX<sup>-</sup></i> W1485	P1(FF1005) $\times$ PF2
IK12	MC4100 $\Phi$ otsB-lacZ18 <i>cys-95::Tn10 otsX<sup>-</sup></i> MC4100	P1(N3002) $\times$ FF1112
IK19	W1485 $\Delta$ lac $\Phi$ otsB-lacZ18 <i>srl-300::Tn10 otsX<sup>-</sup></i> MC4100	P1(1K6) $\times$ IK4
IK21	W1485 $\Delta$ lac $\Phi$ otsB-lacZ18 <i>mutS215::Tn10 otsX<sup>-</sup></i> W1485	P1(ES1481) $\times$ IK4
IK25	W1485 $\Delta$ lac $\Phi$ otsB-lacZ18 <i>cys-95::Tn10 otsX<sup>-</sup></i> MC4100	P1(1K12) $\times$ IK4
IK26	W1485 $\Delta$ lac $\Phi$ otsB-lacZ18 <i>cys-95::Tn10 otsX<sup>-</sup></i> W1485	P1(1K12) $\times$ IK4
IK35	W1485 $\Delta$ lac $\Phi$ otsB-lacZ18 <i>otsX<sup>-</sup></i> LCB107	P1(LCB107) $\times$ IK25
IK41	W1485 $\Delta$ lac $\Phi$ otsB-lacZ18 <i>cys153(Am) otsX<sup>-</sup></i> W1485 <i>supE42</i>	P1(DC943) $\times$ PF2
IK43	MC4100 $\Phi$ otsA-lacZ17 <i>cys-95::Tn10 otsX<sup>-</sup></i> W1485	P1(1K26) $\times$ FF2032
IK44	MC4100 $\Phi$ otsA-lacZ17 <i>otsX<sup>-</sup></i> W1485	P1(FF1112) $\times$ IK43
IK45	MC4100 <i>otsA1::Tn10 otsX<sup>-</sup></i> MC4100	P1(FF4169) $\times$ FF2032
IK46	MC4100 <i>otsA1::Tn10 otsX<sup>-</sup></i> W1485	P1(FF4169) $\times$ IK44
IK49	F <sup>-</sup> 2 $\Phi$ ( <i>betI-lacZ4</i> /MC4100 <i>otsA1::Tn10 otsX<sup>-</sup></i> MC4100	MLE914 $\times$ IK45
IK50	F <sup>-</sup> 2 $\Phi$ ( <i>betI-lacZ4</i> /MC4100 <i>otsA1::Tn10 otsX<sup>-</sup></i> W1485	MLE914 $\times$ IK46
IK51	F <sup>-</sup> 2 $\Phi$ ( <i>betI-lacZ2</i> /MC4100 <i>otsA1::Tn10 otsX<sup>-</sup></i> MC4100	MLE413 $\times$ IK45
IK52	F <sup>-</sup> 2 $\Phi$ ( <i>betI-lacZ2</i> /MC4100 <i>otsA1::Tn10 otsX<sup>-</sup></i> W1485	MLE413 $\times$ IK46
IK58	LCB107 $\Delta$ lac $\Phi$ otsA-lacZ17 <i>metF159(Am) otsX<sup>-</sup></i> LCB107 <i>supE44</i>	P1(BL50) $\times$ IK5
IK59	LCB107 $\Delta$ lac $\Phi$ otsA-lacZ17 <i>metF159(Am) otsX<sup>-</sup></i> LCB107 <i>sup<sup>0</sup> zbf-507::Tn10</i>	P1(JW1071) $\times$ IK58
IK60	LCB107 $\Delta$ lac $\Phi$ otsA-lacZ17 <i>metF159(Am) cysC mutS215::Tn10 otsX<sup>-</sup></i> <i>supE44</i>	P1(1K68) $\times$ IK58
IK61	LCB107 $\Delta$ lac $\Phi$ otsA-lacZ17 <i>metF159(Am) otsX<sup>-</sup></i> MC4100 <i>supE44</i>	P1(FF1112) $\times$ IK60
IK63	LCB107 $\Delta$ lac $\Phi$ otsA-lacZ17 <i>metF159(Am) otsX<sup>-</sup></i> MC4100 <i>sup<sup>0</sup> zbf-507::Tn10</i>	P1(JW1071) $\times$ IK61
IK65 <sup>c</sup>	MC4100 $\Phi$ otsB-lacZ18 <i>rpoS359::Tn10</i>	P1(RH90) $\times$ FF1112
IK68	F <sup>-</sup> <i>cysC argA lysA pro mutS215::Tn10 otsX<sup>-</sup></i>	P1(ES1481) $\times$ KM78
IK69 <sup>c</sup>	W1485 $\Delta$ lac $\Phi$ otsB-lacZ18 <i>rpoS359::Tn10</i>	P1(RH90) $\times$ IK4
IK70 <sup>c</sup>	W1485 $\Delta$ lac $\Phi$ otsB-lacZ18 <i>cys153(Am)? rpoS359::Tn10 supE42</i>	P1(RH90) $\times$ IK41
JC10240	Hfr <i>recA56 thr-300 srl-300::Tn10 thi-1 relA1 ilv-318 spoT1 rpsE2300 <math>\lambda^-</math> otsX</i>	CGSC 6074
JW1071	Hfr <i>sup<sup>0</sup> zbf-507::Tn10 irp-49 lacZ125 relA1 spoT1 <math>\lambda^-</math></i>	CGSC 6391
K-12	F <sup>-</sup> wild type	
KM78	F <sup>-</sup> <i>cysC argA lysA pro</i>	W. Epstein
LCB107	F <sup>-</sup> <i>ire-1 trpA43 metB1 lacY1 malA1 (<math>\lambda^-</math>) rpsL134 otsX<sup>-</sup></i> LCB107 <i>supE44?</i>	CGSC 6407 (6)
MC4100	F <sup>-</sup> <i>araD139 <math>\Delta</math>(argF-lac)U169 ffbB5301 relA1 rpsL150 deoC1 pifF25 rbsR otsX<sup>-</sup></i> MC4100	CGSC 6152
MLE413	F <sup>-</sup> 2 $\Phi$ ( <i>betI-lacZ2</i> /MC4100 <i>recA56 Rif</i>	M. W. Eshoo (17)

Continued on following page

TABLE 1—Continued

Strain, bacteriophage, or plasmid	Description <sup>a</sup>	Construction or source <sup>a</sup>
MLE914	F <sup>+</sup> $\Phi$ hetT-lacZ/MC4100 recA56 Rif <sup>r</sup>	M. W. Esboo (17)
N3002	F <sup>+</sup> cys-95::Tn10 IN(rrnD-rrnE)1 $\lambda$ <sup>+</sup> <i>otsX</i>	CGSC 646
OS18	W1485 $\Delta$ (argF-lac)U169 zah-735::Tn10 <i>otsX</i> <sub>W1485</sub>	P1(SH202) $\times$ W1485
OS19	W1485 $\Delta$ (argF-lac)U169 zah-735::Tn10 <i>cysI53</i> (Am) <i>lacZ125</i> (Am) <i>otsX</i> <sub>W1485</sub>	P1(SH202) $\times$ DC931
OS36	LCB107 $\Delta$ (argF-lac)U169 zah-735::Tn10 $\Phi$ otsA-lacZ17 <i>otsX</i> <sub>LCB107</sub> <i>supE44</i>	52
PF1	W1485 $\Delta$ lac <i>cysI53</i> (Am) <i>lacZ125</i> (Am) <i>otsX</i> <sub>W1485</sub>	Tc <sup>r</sup> of OS19
PF2	W1485 $\Delta$ lac $\Phi$ otsB-lacZ18 <i>cysI53</i> (Am) <i>otsX</i> <sub>W1485</sub>	P1(FF1123) $\times$ PF1
PF11	BL50 <i>zih-730</i> ::Tn10	P1(GD1) $\times$ BL50
RH90 <sup>r</sup>	MC4100 <i>rpoS359</i> ::Tn10	R. Hengge-Aronis (30)
SH205	HfrC <i>phoA8 glpD3 glpR2 relA1 tonA22</i> ( $\lambda$ ) $\Delta$ (argF-lac)U169 zah-735::Tn10	E. Bremer (48)
UES	F <sup>+</sup> KL16 <i>thi</i> $\Delta$ (ptsHI-crr) <i>galR</i> <i>trcA</i> ::Tn10	8
W1485	F <sup>+</sup> wild type <i>otsX</i> <sub>W1485</sub>	CGSC 504
Phages		38
P1	<i>cml clr-100</i>	E. Bremer (49)
$\lambda$ NK561	b221 <i>cl</i> ::Tn10 <i>Oam29</i> <i>Pam80</i>	
Plasmids		46
cos4	Ap <sup>r</sup>	12
pACYC184	Tc <sup>r</sup> Cm <sup>r</sup>	This study
pFF1	Ap <sup>r</sup> ; vector cos4	This study
pFF101	Ap <sup>r</sup> ; vector cos4	This study
pFF102	Ap <sup>r</sup> ; vector cos4	This study
pFF106	Cm <sup>r</sup> ; vector pACYC184	This study
pFF109	Ap <sup>r</sup> ; vector cos4	This study
pFF114	Cm <sup>r</sup> ; vector pACYC184	This study

<sup>a</sup> *otsX* is allelic with *kaiF*, *appR*, *csi-2*, and *rpoS*. *otsX* carries an amber mutation in many K-12 strains (see text). The *otsX* genotype is shown where it is relevant, and when known, the origin of the *otsX* allele is indicated by a subscript. All strains described as  $\Delta$ lac carry a deletion generated by selecting a Tc<sup>r</sup> derivative from a strain carrying  $\Delta$ (argF-lac)U169 zah-735::Tn10. The  $\Delta$ (argF-lac)U169 deletion is 100% cotransducible with the transposon (48). The symbol  $\Phi$  indicates that the strain carries a *lacZ* operon fusion generated by a  $\lambda$  *placMu53* or a *placMu55* insertion.

<sup>b</sup> CGSC, *E. coli* Genetic Stock Center, Yale University, New Haven, Conn. All CGSC strains were obtained from B. J. Bachmann.

<sup>c</sup> These strains carry a  $\lambda$  *placMu55* insertion in which the kanamycin resistance marker of the prophage is deleted together with *otsA*::Tn10.

<sup>d</sup> We do not know whether the  $\Delta$ (cys::Tn10) deletion encompasses the *otsX*<sub>W1485</sub> allele.

<sup>e</sup> The *kaiF* allele *rpoS359*::Tn10 is identical with *csi-2*::Tn10. This mutation was generated in MC4100 (30).

assayed on agar plates (30). Trehalose-6-phosphate was a gift from Marine DNA, Tromsø, Norway.

## RESULTS

**Identification of *otsX*.** Most suppressor-free K-12 strains (e.g., *sup*<sup>0</sup> derivatives of W1485) display a lower expression of *otsA-lacZ* and *otsB-lacZ* operon fusions than their counterparts with an amber suppressor (i.e., *supD*, *supE*, or *supF*). However, the much-used *sup*<sup>0</sup>-containing strain MC4100 is an exception; its derivatives display the same high expression of *ots-lacZ* fusions as strains containing *sup*<sup>+</sup> (52). To make certain that our strain of MC4100 did not display a high expression of the *otsA* and *otsB* genes because of an unknown amber suppressor, we verified that a *metF*(Am) mutation was not suppressed in this genetic background (see Materials and Methods).

In order to insert Tn10 near the gene which activated *otsA* and *otsB* in MC4100, we prepared a culture of MC4100 with random Tn10 insertions. A P1 lysate grown on this culture was then used to transduce PF2 [W1485 *sup*<sup>0</sup> *otsB-lacZ* *cysI*(Am)], which displayed a low expression of its *lacZ* fusion, to Tc<sup>r</sup>. The Tc<sup>r</sup> transductants were scored for growth on plates with medium 63-lactose. On this agar medium, strains with high or low expression of *lacZ* fusions in *otsA* or *otsB* are known to display a Lac<sup>+</sup> or Lac<sup>-</sup> phenotype, respectively (52). Out of 800 Tc<sup>r</sup> transductants tested, 6 were Lac<sup>+</sup>.

Our rationale for using PF2 [*cysI*(Am)] as the recipient in this cross was to detect whether a Lac<sup>+</sup> phenotype of the transductants was caused by a spontaneous amber suppressor mutation. But fortuitously, all six Tc<sup>r</sup> Lac<sup>-</sup> transductants selected appeared to be Cys<sup>-</sup>. When P1 lysates prepared from these Tc<sup>r</sup> Cys<sup>+</sup> transductants were used to transduce DC931 [*cysI*(Am) *lacZ*(Am)] to Tc<sup>r</sup>, we obtained Tc<sup>r</sup> Cys<sup>+</sup> transductants at a high frequency, but none was Tc<sup>r</sup> Lac<sup>+</sup>. Furthermore, when the same lysates were used to transduce AT2427, which carries a *cysI* point mutation, we also obtained Tc<sup>r</sup> Cys<sup>+</sup> transductants. Thus, these crosses established that the gene in MC4100 that had the ability to provoke a high expression of the *lacZ* fusions in a *sup*<sup>0</sup> background maps near *cysI* at 59 min. We tentatively named this gene *otsX*, and for clarity we name the MC4100 allele *otsX*<sup>+</sup><sub>MC4100</sub> and the W1485 allele *otsX*<sub>W1485</sub>. In other words, alleles that caused elevated transcription of the *otsA* and *otsB* genes in a *sup*<sup>0</sup> background are designated *otsX*<sup>+</sup>.

***otsX* genotype and strain construction.** In order to know the origin of the *otsX* allele of the constructed strains, we first checked the *otsX* genotype of the basic strains. This was done by transducing their *otsX* allele together with a neighbor selectable marker into two recipient strains, one carrying the *otsX*<sup>+</sup><sub>MC4100</sub> allele and the other the *otsX*<sub>W1485</sub> allele. Both recipients were *sup*<sup>0</sup>  $\Delta$ lac and carried either *otsA-lacZ* or *otsB-lacZ*. When the donor and recipient had the same *otsX* genotype, all transductants had the same Lac phenotype as the recipient. When the donor and the recipient had

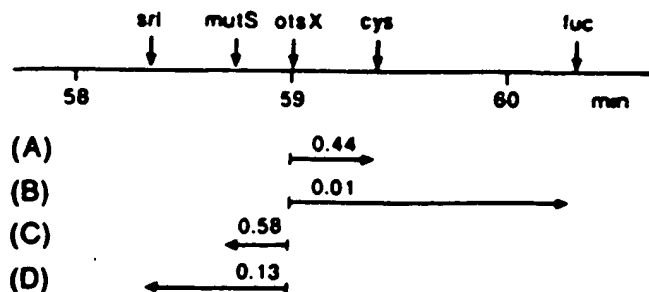


FIG. 1. Mapping of *otsX* by cotransductional analysis. The values given on the arrows show the P1 cotransduction frequencies between *otsX* and various markers in the 59-min region. The arrows point at the selected marker, which in all cases was a Tn10 insertion. All recipients were  $\Delta$ lac and carried an *otsB-lacZ* operon fusion, and the *otsX* genotype of the Tc<sup>r</sup> transductants was scored by determining their Lac phenotype on agar plates with medium 63-lactose. From each cross, at least 100 Tc<sup>r</sup> transductants were screened. Cross A, donor IK12 (*cys::Tn10 otsX*<sup>+</sup> MC4100) and recipient IK4 (*otsX*<sup>+</sup> W1485); cross B, donor IK1 (*fuc::Tn10 otsX*<sup>+</sup> MC4100) and recipient PF2 (*otsX*<sup>+</sup> W1485); cross C, donor IK21 (*mutS::Tn10 otsX*<sup>+</sup> W1485) and recipient FF1112 (*otsX*<sup>+</sup> MC4100); cross D, donor IK6 (*srl::Tn10 otsX*<sup>+</sup> MC4100) and recipient IK4 (*otsX*<sup>+</sup> W1485).

differing types of *otsX* alleles, the transductants would be of both Lac phenotypes. By setting up crosses in which the *otsX* genotype of the donor and the recipient differed, we could always pick a progeny with the desired *otsX* allele. This strategy was therefore used in the strain constructions listed in Table 1. Beside MC4100, we found that ES1481 carries an *otsX*<sup>+</sup> allele, whereas JC10240, N3002, and DC906 are *otsX*<sup>-</sup>.

*otsX*<sup>+</sup>-containing strains that also were *otsA*<sup>+</sup> *otsB*<sup>+</sup> could be distinguished from isogenic *otsX* mutants by their osmotically somewhat more tolerant phenotype, but this phenotype was not strong enough to be used for positive selection in a transductional cross. We also avoided a direct selection for a Lac<sup>+</sup> phenotype of *otsA-lacZ* or *otsB-lacZ* mutants, because strains carrying a fusion together with an *otsX* allele gave rise to spontaneous Lac<sup>+</sup> mutants when streaked on agar medium with lactose as an energy source. Ten such mutations, examined by transductional analysis, appeared to be linked to the fusion; thus, they seemed to be promoter mutations (data not shown).

**Transductional mapping of *otsX*.** We constructed strains which carried *fuc::Tn10*, *cysC::Tn10*, *mutS::Tn10*, or *srl::Tn10* together with either the *otsX*<sup>+</sup> MC4100 or the *otsX*<sup>+</sup> W1485 allele. These strains were used as donors in transductional crosses in which the recipient carried an *otsX* allele of the opposite category and an *otsB-lacZ* fusion. The transductants were selected for the Tn10 marker (i.e., Tc<sup>r</sup>) and the cotransduction frequency of *otsX* and Tn10 was determined by scoring the Lac phenotype. The four crosses outlined in Fig. 1 placed *otsX* almost exactly at 59 min.

In a three-factor cross, a P1 lysate prepared from strain IK68 (*otsX*<sup>+</sup> *mutS::Tn10* *cysC*) was used to transduce strain IK4 (*otsX*<sup>+</sup> W1485 *otsB-lacZ*) to Tc<sup>r</sup>. Of 175 transductants tested, 54 had received only *mutS::Tn10*, 89 had received *mutS::Tn10* and *otsX*<sup>+</sup>, and 32 had received *mutS::Tn10*, *otsX*<sup>+</sup>, and *cysC*. The lack of transductants which had received *mutS::Tn10* and *cysC*, but not *otsX*<sup>+</sup>, showed that the gene order was *mutS-otsX-cysC*. This is in accordance with data presented in Fig. 1.

In addition to *otsX*, *katF* (32), *appR* (53), and *csi-2* (*rpoS*) (30) have been mapped to 59 min (4). It has been shown that

TABLE 2. Influence of the *otsX* (*katF*) allele and genetic background on  $\beta$ -galactosidase expression from *otsA-lacZ* and *otsB-lacZ* operon fusions

Strain	Description <sup>a</sup>	$\beta$ -Galactosidase activity (U/mg of protein) <sup>b</sup> when grown in:	
		LOM	LOM + 0.3 M NaCl
IK3	W1485 <i>sup</i> <sup>o</sup> <i>otsA-lacZ</i> <i>otsX</i> <sup>+</sup> W1485	20	62
IK4	W1485 <i>sup</i> <sup>o</sup> <i>otsB-lacZ</i> <i>otsX</i> <sup>+</sup> W1485	10	44
IK41	W1485 <i>supE</i> <i>otsB-lacZ</i> <i>otsX</i> <sup>+</sup> W1485	77	780
IK19	W1485 <i>sup</i> <sup>o</sup> <i>otsB-lacZ</i> <i>otsX</i> <sup>+</sup> MC4100	120	850
IK35	W1485 <i>sup</i> <sup>o</sup> <i>otsB-lacZ</i> <i>otsX</i> <sup>+</sup> LCB107	9	39
FF1112	MC4100 <i>sup</i> <sup>o</sup> <i>otsB-lacZ</i> <i>otsX</i> <sup>+</sup> MC4100	130	540
IK44	MC4100 <i>sup</i> <sup>o</sup> <i>otsA-lacZ</i> <i>otsX</i> <sup>+</sup> W1485	18	57
IK58	LCB107 <i>supE</i> <i>otsA-lacZ</i> <i>otsX</i> <sup>+</sup> LCB107	360	1,500
IK59	LCB107 <i>sup</i> <sup>o</sup> <i>otsA-lacZ</i> <i>otsX</i> <sup>+</sup> LCB107	33	89
IK63	LCB107 <i>sup</i> <sup>o</sup> <i>otsA-lacZ</i> <i>otsX</i> <sup>+</sup> MC4100	450	1,800
IK65	MC4100 <i>sup</i> <sup>o</sup> <i>otsB-lacZ</i> <i>katF::Tn10</i>	5	18
IK69	W1485 <i>sup</i> <sup>o</sup> <i>otsB-lacZ</i> <i>katF::Tn10</i>	7	21
IK70	W1485 <i>supE</i> <i>otsB-lacZ</i> <i>katF::Tn10</i>	8	22
MC4100		0	0

<sup>a</sup> All strains are  $\Delta$ lac.

<sup>b</sup> One unit of enzyme activity is 1  $\mu$ mol of *o*-nitrophenol formed per min at 28°C. Each value is an average of at least three independent measurements. Standard deviations were within  $\pm 15\%$ .

*katF* is allelic with *appR* (54), as well as with *csi-2* (*rpoS*) (30), and our results show that *otsX* is another allele of this locus (see below).

**Influence of the *otsX* allele and the genetic background on *otsA* and *otsB* expression.** To measure quantitatively the influence of the *otsX* alleles on the expression of *otsA* and *otsB*, we assayed the  $\beta$ -galactosidase activity of *lacZ* fusion-containing strains with known genetic backgrounds. The activity was assayed after growth in a medium with or without 0.3 M NaCl added. *otsA-lacZ* or *otsB-lacZ* fusions in the same genetic background displayed similar values for  $\beta$ -galactosidase (19, 52). Therefore, in most cases only one type of fusion is included for each combination of genetic background and *otsX* allele (Table 2).

In accordance with data presented before (52), strain IK4, which is a suppressor-free strain with a W1485 background and the native *otsX*<sup>+</sup> W1485 allele, displayed low  $\beta$ -galactosidase activity. Furthermore, the activity was much higher in W1485 derivatives that carried either a *supE* mutation or the *otsX*<sup>+</sup> MC4100 allele, i.e., IK41 and IK19, respectively (Table 2).

The difference between W1485 and MC4100 derivatives with respect to their expression of *otsA-lacZ* and *otsB-lacZ* fusions was solely due to the *otsX* alleles, since the genetic background, MC4100 or W1485, was of no importance when the *otsX* allele remained the same. Thus, IK44 (MC4100 *otsX*<sup>+</sup> W1485) displayed a  $\beta$ -galactosidase activity similar to that of IK3 (W1485 *otsX*<sup>+</sup> W1485), and FF1112 (MC4100 *otsX*<sup>+</sup> MC4100) displayed an activity similar to that of IK19 (W1485 *otsX*<sup>+</sup> MC4100) (Table 2).

We have reported previously that LCB107 derivatives display a higher expression of *otsA* and *otsB* than MC4100 derivatives, and we have concluded that this property is not due to the *otsA* and *otsB* genes themselves (52). The *otsX* allele of LCB107 was not the cause of the elevated expression of the *otsA* and *otsB* genes, since the isogenic *otsB-lacZ* fusion-containing strains IK4 (W1485 *otsX*<sup>+</sup> W1485) and IK35

(W1485 *otsX*<sub>LCB107</sub>) displayed similar low  $\beta$ -galactosidase levels.

From the construction of LCB107, it was uncertain whether it carried *supE44* (5). Our study showed that LCB107 carries a *supE* mutation (see Materials and Methods). The *supE*-containing strain IK59, derived from LCB107, displayed strongly reduced expression of the *otsA-lacZ* fusion compared with that of the parental *supE*-containing strain IK58. However, *supE* was not responsible for the superexpression of the *otsA* and *otsB* genes in LCB107 (i.e., expression above the MC4100 level), since IK63 (LCB107 *supE* *otsX*<sup>+</sup> MC4100) expressed the *otsA-lacZ* fusion at the same very high level as IK58 (LCB107 *supE* *otsX*<sub>LCB107</sub>). Apparently, LCB107 carries a mutation in an as yet unidentified gene which influences the expression of the *otsA* and *otsB* genes.

*otsX* is allelic with *katF* and the amber-mutated gene of *E. coli* K-12. We found that all *supE*-containing strains listed in Table 2 carrying the *otsX*<sub>W1485</sub> or the *otsX*<sub>LCB107</sub> allele, and also *E. coli* K-12 itself, displayed the glycogen-negative and catalase-negative phenotype that is found in *katF* mutants (30), while all strains containing *otsX*<sup>+</sup> MC4100 or *supE* did not show these defects. Also, when a *katF*::Tn10 mutation was introduced into FF112 (MC4100 *supE* *otsX*<sup>+</sup> MC4100), the resulting strain, IK65, did not display the high expression of the *lacZ* fusion seen in the parental strain. In fact, the  $\beta$ -galactosidase activities found were even lower than that of IK4 (W1485 *supE* *otsX*<sub>W1485</sub>), indicating that the *katF*::Tn10 mutation reduced *otsA* and *otsB* gene transcription even more than *otsX*<sub>W1485</sub> (Table 2). The simplest explanation of the present data is that *otsX*, *katF*, and the amber mutation described previously (47, 52; see above) are allelic and that this gene has reverted to an active form in MC4100.

An alternative explanation is that the amber mutation was situated in another locus and was suppressed by the *otsX*<sup>+</sup> (*katF*<sup>+</sup>) allele of MC4100. If so, the *otsX*<sup>+</sup> MC4100 allele and tRNA amber suppressors would represent two different routes for intergenic suppression of the amber mutation. However, the latter explanation was ruled out by the finding that IK70 (W1485 *supE* *katF*::Tn10) expressed the *otsA-lacZ* fusion at the same low level as IK69 (W1485 *supE* *katF*::Tn10) (Table 2). Apparently, in these strains the *katF*::Tn10 allele had replaced the amber-mutated gene so that suppression via *supE* was no longer possible.

It is well known that many K-12 strains lose their viability rather rapidly when stored as colonies on LB agar in the cold. We found that this phenotype was linked to the *otsX* (*katF*) genotype. This loss of viability was not caused by decreased trehalose synthesis, since *otsX*<sup>+</sup> (*katF*<sup>+</sup>) strains carrying an *otsA* or *otsB* mutation remained viable for a prolonged period (data not shown).

*otsX* (*katF*) does not activate *proU* and *bet*. To investigate the influence of *otsX* on the transcription of the osmotically regulated *betB*, *betT* (17), and *proU* (15) genes, we used suppressor-free strains which were blocked in trehalose synthesis by an *otsA*::Tn10 mutation (19). Thereby we could eliminate any indirect effect of *otsX* mediated via trehalose synthesis. As shown in Table 3, the  $\beta$ -galactosidase activities of cells with a *betB-lacZ* or *betT-lacZ* operon fusion were essentially the same whether the cells carried the *otsX*<sup>+</sup> MC4100 or the *otsX*<sub>W1485</sub> allele. The same lack of effect of *otsX* was obtained for strains carrying a *proU-lacZ* fusion (data not presented). Thus, *otsX* is not a general activator of osmotically regulated genes.

Cloning of *otsA* and *otsB*. Strains which are defective in trehalose synthesis display an osmotically sensitive pheno-

TABLE 3. Influence of *otsX* (*katF*) allele on the osmotic induction of  $\beta$ -galactosidase activity of *betT-lacZ* and *betB-lacZ* operon fusion mutants of *E. coli*

Strain	Description	$\beta$ -Galactosidase activity (1 ng of protein) <sup>a</sup> of cells grown in:	
		LOM	LOM + 0.3 M NaCl
IK49	<i>otsX</i> <sup>+</sup> MC4100 <i>betT-lacZ</i>	150	600
IK50	<i>otsX</i> <sub>W1485</sub> <i>betT-lacZ</i>	160	670
IK51	<i>otsX</i> <sup>+</sup> MC4100 <i>betB-lacZ</i>	350	670
IK52	<i>otsX</i> <sub>W1485</sub> <i>betB-lacZ</i>	320	750

<sup>a</sup> All strains are  $\Delta$ *lac*.

<sup>b</sup> One unit of enzyme activity is 1 nmol of o-nitrophenol formed per min at 28°C. Each value is an average of at least three independent measurements. Standard deviations were within  $\pm 15\%$ .

type (19). By introducing a cosmid library into FF4169 (*otsA*) and selecting for strains growing on agar with medium 63-glucose with 0.45 M NaCl added, we obtained plasmid pFF1. This plasmid carried a chromosomal insert of 43 kb (Fig. 2). From pFF1 we constructed plasmid pFF106 carrying a 2.9-kb *Hind*III chromosomal fragment. This was the smallest fragment obtained that restored trehalose synthesis in both *otsA* and *otsB* mutants, as well as in  $\Delta$ (*otsA otsB*) mutants. We also constructed the subclones pFF109 and pFF114, which carried overlapping chromosomal fragments with lengths of 1.85 and 2.1 kb, respectively, covering the

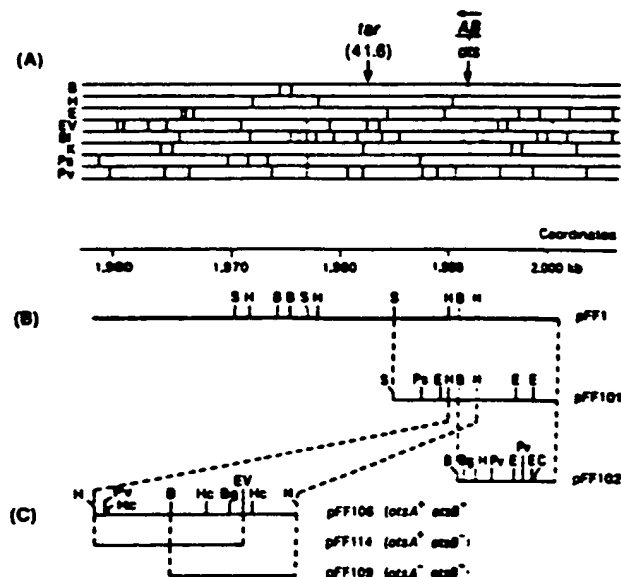


FIG. 2. Restriction site analysis of the *otsBA* region. (A) Restriction map of the 41- to 42-min region of the *E. coli* chromosome as published by Kohara et al. (28); the coordinates are in kilobases from *thr*. The localization and the direction of transcription of the *otsBA* operon are indicated. (B) Restriction map of the chromosomal part of plasmid pFF1 and its subclones pFF101 and pFF102, aligned with the map of Kohara et al. (C) Enlarged restriction map of the 2.9-kb *Hind*III fragment containing the *otsBA* operon. The extensions of the chromosomal part of plasmid pFF106, pFF109, and pFF114 are shown, and their genotypes are indicated. Abbreviations used for restriction sites are as follows: B, *Bam*HI; Bl, *Bgl*II; Bg, *Bgl*II; C, *Cla*I; E, *Eco*RI; EV, *Eco*RV; H, *Hind*III; Hc, *Hinc*II; K, *Kpn*I; Ps, *Pst*I; Pv, *Pvu*II; and S, *Sal*I.

TABLE 4. Osmotic regulation of the trehalose-6-phosphate synthase and the trehalose-6-phosphate phosphatase activities in *E. coli* strains with and without *ots* mutations and plasmids

Strain <sup>a</sup>	Synthase activity (U/mg of protein) <sup>b</sup> in cells exposed to:		Phosphatase activity (U/mg of protein) <sup>b</sup> in cells exposed to:	
	M63	M63 + 0.4 M NaCl	M63	M63 + 0.4 M NaCl
FF4171	2	17	22	88
FF4026 ( <i>otsA-lacZ</i> )	0	0	28	100
FF4025 ( <i>otsB-lacZ</i> )	0	0	28	28
FF4050 ( $\Delta$ ( <i>otsA otsB</i> ))	0	0	22	22
FF4050(pFF106)	26	170	66	200
FF4050(pFF114)	11	15-90 <sup>c</sup>	14	14
FF4050(pFF109)	0	0	71	530
FF4057(pFF106) ( <i>otsX</i> )	ND	27	ND	50

<sup>a</sup> All strains were derived from MC4100. *otsX* is allelic with *katF*.

<sup>b</sup> Enzyme activities were measured in toluene-treated cells. One unit of enzyme activity is 1  $\mu$ mol of trehalose formed per min at 37°C. Each value is an average of at least four independent measurements. Standard deviations were within  $\pm 10\%$  for the synthase activities, except for FF4050(pFF114) (see text), and within  $\pm 20\%$  for the phosphatase activities. ND, not determined.

<sup>c</sup> The synthase activity displayed large variations (see text).

insert of pFF106 (Fig. 2). Properties of these three plasmids are described below.

Expression of synthase and phosphatase activities. We previously described an assay for trehalose-6-phosphate synthase in *E. coli* (19, 52). In this work we have developed an assay for the phosphatase of the trehalose pathway, using synthetic trehalose-6-phosphate as a substrate and determining, by using gas chromatography, the amount of trehalose formed. In order to measure the osmotic induction of the enzymes, we grew the cells in medium 63 and then increased the osmolarity of the medium by addition of 0.4 M NaCl. Strains with intact trehalose synthesis which grew in the presence of 0.4 M NaCl were incubated for one generation, whereas strains with a defective trehalose synthesis were incubated in the high-osmolarity medium for 2 h. All strains used were derivatives of MC4100 and carried a *trcA* mutation to prevent in vitro degradation of trehalose (19). Except for FF4057, they were all *otsX*<sup>+</sup> (*katF*<sup>+</sup>).

Strain FF4171 is wild type with respect to trehalose synthesis. As shown previously (19, 52), FF4171 has an osmotically inducible synthase, whereas related strains carrying an *otsA*, *otsB*, or  $\Delta$ (*otsA otsB*) mutation lack this synthase activity completely (19) (Table 4). In addition, FF4171 displayed an osmotically induced phosphatase activity. The phosphatase activity was also osmotically induced in FF4026 (*otsA*), but not in FF4025 (*otsB*) and FF4050 ( $\Delta$ (*otsA otsB*)), in which it was the same whether the cells were osmotically stressed or not (Table 4). Apparently, the *otsB* mutation blocked the synthesis of the osmotically inducible phosphatase, but it did not influence the background phosphatase activity also found in nonstressed cells. It is well known that it is often difficult to measure accurately separate phosphatase activities in wild-type *E. coli*. This is because of the multiplicity of phosphatases in *E. coli* and the lack of specificity of some of them (42).

Strain FF4050 ( $\Delta$ (*otsA otsB*)) carrying plasmid pFF106 displayed strongly elevated activities of both the synthase and the phosphatase. Both activities, after subtraction of the background phosphatase activity, increased about fivefold when FF4050(pFF106) was exposed to osmotic stress. The synthase and phosphatase activities of strains carrying plas-

mid pFF106 (*otsA*<sup>+</sup> *otsB*<sup>+</sup>) were much lower in the *otsX*<sup>-</sup> (*katF*<sup>-</sup>) mutant strain FF4057 than in its *otsX*<sup>+</sup> counterpart FF4050 (Table 4).

Subclone pFF114, which restored trehalose synthesis in synthase-defective *otsA* mutants, conferred only synthase activity to the  $\Delta$ (*otsA otsB*) mutant FF4050. Subclone pFF109 conferred only phosphatase activity to FF4050 (Table 4). As would be expected, pFF109 did not restore trehalose synthesis in either *otsA* or *otsB* insertion mutants, since these mutants lacked the synthase activity. The elevated enzymic activities displayed by the present plasmid-carrying strains and the finding that the enzymes can be expressed separately indicated that the plasmids carried structural rather than regulatory genes. Presumably, *otsA* and *otsB* were the structural genes of the synthase and the phosphatase, respectively, and the lack of synthase activity in *otsB* mutants was because of a polar effect.

Difference in copy number may explain why pFF109 (pBR322 derived) appeared to confer a much higher phosphatase activity to osmotically stressed cells than did pFF106 (pACYC derived) (Table 4). The lower ratio of osmotic induction of the phosphatase seen with pFF106 compared with pFF109 may be because of a constitutive contribution to *otsBA* transcription from a vector promoter in pFF106. Stüber and Bujard (51) have shown that pACYC184 contains a strong promoter situated just upstream of the *otsBA* insert in pFF106.

Addition of spectinomycin (300  $\mu$ g ml<sup>-1</sup>) reduced strongly the osmotic induction of the enzyme activities in all plasmid-carrying strains (data not shown). Thus, the observed osmotic effect depended on de novo protein synthesis. It should, however, be noted that the synthase activity of stressed cells of FF4050(pFF114) was very variable (Table 4). This was probably due to a toxic effect of intracellular trehalose-6-phosphate, which accumulated in the absence of the specific phosphatase activity. This notion is supported by the finding that when FF4050(pFF114) was subjected to an osmotic upshock, the survival of the cells, as determined by counting CFU on LB plates, was only 50% after 2 h and 10% after 4 h. It has been reported previously that intracellular trehalose-6-phosphate is inhibitory to both *S. cerevisiae* (41) and *E. coli* (27).

Physical mapping of *otsBA*. The 43-kb chromosomal fragment of pFF1 was mapped with *Bam*HI, *Hind*III, and *Sal*I. The subclone pFF101 was in addition mapped with *Pst*I and *Eco*RI, and the shorter subclone pFF102 also with *Bgl*II, *Pvu*II, and *Cla*I. In Fig. 2, the restriction maps of the chromosomal fragments of pFF1 and its subclones have been aligned with the 41- to 42-min region of the physical map of the *E. coli* chromosome prepared by Kohara et al. (28). A very good agreement was found for all enzymes used by both Kohara et al. and us. The only exception was the *Hind*III site furthest to the right in pFF1, which was not reported by Kohara et al. This is the *Hind*III site limiting the right-hand side of the chromosomal insert of pFF106 (Fig. 2).

In conclusion, the 2.9-kb *Hind*III fragment which complements  $\Delta$ (*otsA otsB*) mutants is localized around the coordinate 1,992 kb in the physical map of the *E. coli* chromosome (28). That is 8 to 11 kb clockwise of *tar* (41.6 min). On the basis of the localization of the chromosomal fragment of subclones pFF109, conferring only OtsB (phosphatase) activity, and pFF114, conferring only OtsA (synthase) activity, it was evident that *otsA* is located proximal to *tar* with respect to *otsB*. This localization and gene order is in agreement with our previous mapping data (19). However, on the latest chromosomal linkage map of *E. coli* these *ots*



genes are wrongly placed between *flaA* and *uvrC* (4) and not in the correct position between *uvrC* and *flhD* (*flbB*) (19).

It has been shown previously that a Tn10 insertion in *otsA* does not influence the expression of a *lacZ* fusion in *otsB* (19). However, the present biochemical data showed that an insertion in *otsB* prevented the expression of *otsA* (Table 4). This polar effect suggests that *otsBA* constitutes an operon. The operon model is also supported by the finding that when preparing deletion mutants from FF4012 (MC4100 *otsA*::Tn10 *otsB-lacZ otsX*<sup>MC4100</sup>) by selecting for Tc<sup>r</sup> and scoring for the Km Lac phenotype, we obtained a few strains, e.g., strain FF4035, which were Tc<sup>r</sup> Km<sup>r</sup> but remained Lac<sup>-</sup>. The *otsB-lacZ* fusion in FF4035 remained under the control of the native *otsX* (*kaiF*)-dependent promoter since it became Lac<sup>-</sup> when an *otsX*<sub>W1485</sub> allele was introduced to generate strain FF4055. Apparently, the deletion of *otsA*::Tn10 from FF4012 had been accompanied by a deletion of the kanamycin resistance marker, but not the *lacZ* gene of the  $\lambda$  *placMu* insertion. The kanamycin marker is located distal to the promoter controlling the expression of *lacZ* in the intact  $\lambda$  *placMu* prophage. Thus, these data showed that the *otsB* gene is transcribed in the direction towards *otsA*, i.e., in the counterclockwise direction on the chromosome. We assume that the transcription of *otsA* in pFF114 (Table 4) starts from a promoter which is not functional on the chromosome.

## DISCUSSION

The *otsA* and *otsB* genes have been mapped previously (19). In this study we have cloned these genes and determined their localization on the physical map of the *E. coli* chromosome prepared by Kohara et al. (28). We present evidence that *otsB* and *otsA* constitute an operon in which *otsB*, encoding trehalose-6-phosphate phosphatase, is proximal to the promoter, and *otsA*, encoding trehalose-6-phosphate synthase, is distal. Thus, a recently described point mutation named *otsP*, which maps near *otsA* and seems to influence only the expression of the phosphatase, is probably allelic with *otsB* (27).

For *S. cerevisiae*, the synthase and the phosphatase of the trehalose pathway have been copurified as a complex and the individual enzymic activities have not yet been resolved (33). This enzyme complex may not exist in *E. coli*, since the *OtsA* and *OtsB* activities could be expressed separately in an *otsBA* deletion mutant. Also supporting this is the finding that we were unable to regain the synthase activity from cells disrupted in a French pressure cell, while the phosphatase activity remained stable (data not shown).

It has been shown previously that the variation among K-12 strains with respect to their capacity to synthesize trehalose and to express *otsBA* is because of an amber-mutated gene in *E. coli* K-12 and many of its derivatives (47, 52). In this study we have identified this mutated gene as *kaiF*, which has recently been reported to be allelic with *appR* (54) and *csi-2* (*rpoS*) (30). The DNA sequence indicates that it encodes a sigma factor (39). Our demonstration that *kaiF* carries an amber mutation in *E. coli* K-12 and W1485 (wild type), but not in MC4100, is in agreement with earlier reports that *appR* is inactive in many K-12 strains (16, 53) and that MC4100 is among the strains in which *appR* is active (16).

Genes which display increased expression when the cell is exposed to a particular external stimulus or stress are often grouped into stimulons. A stimulon can comprise several regulons, and many genes belong to more than one stimulon.

For instance, carbon starvation (37) and osmotic stress (10) are reported to increase the synthesis of at least 54 and 41 proteins, respectively, and 5 osmotic-stress proteins are reported to be among the starvation proteins (25). In experiments with strain MC4100, it has been shown that *kaiF* positively controls 15 to 20 genes expressed during carbon-starvation-induced entry into stationary phase (30). However, the characterization of stimulons in *E. coli* has been carried out with different strains. Thus, the fact that *E. coli* K-12 and many of its derivatives carry an amber mutation in *kaiF* may have made investigators overlook *kaiF*-dependent induction. This has to be taken into account in further evaluations of *E. coli* stimulons.

The *kaiF* gene is known to be transcriptionally regulated, and it has been suggested that slow growth is a signal for its transcription (30, 40). Since osmotic stress slows the growth of *E. coli*, it probably causes *kaiF* induction, and this may partly explain the observed osmotic regulation of *otsBA* transcription. However, it is unlikely that this is the only mechanism, since the *otsBA* transcription, albeit being low, was osmotically regulated also in strains carrying *kaiF*::Tn10 or the amber-mutated *kaiF* gene.

*proU* transcription is dependent on the housekeeping sigma factor  $\sigma^{70}$  (26), and the osmotic regulation of *proU* can be explained by a stimulatory effect of potassium glutamate on the transcriptional complex (43), independent of DNA supercoiling (44). Apparently, an intrinsic property of the osmotic stress-dependent *proU* promoter elicits this potassium glutamate effect without the need of protein factors, at least in vitro (43). A similar effect of potassium glutamate on the *otsBA* promoter could explain the osmotic induction of *otsBA* seen in strains mutated in *kaiF*.

Trehalose protects organisms against a variety of stresses (see the introduction). The finding that transcription of *otsBA*, as opposed to *proU* and *bet*, is *kaiF* dependent nourishes a notion that trehalose, in addition to being an osmoprotectant, may serve as a more general stress protectant in *E. coli*.

## ACKNOWLEDGMENTS

We thank W. Boos and R. Hengge-Aronis for useful discussions. We are grateful to B. J. Bachmann, E. Bremer, D. P. Clark, W. Epstein, M. W. Eshoo, R. Hengge-Aronis, and B. Lowsky for providing bacterial strains and to P. I. Larsen and P. Kolsaker at Marine DNA, Tromsø, Norway, for providing trehalose-6-phosphate.

This work was supported by a grant from the Norwegian Fisheries Research Council.

## REFERENCES

- Andresen, P. A., I. Kaasen, O. B. Styrvold, G. Boulnois, and A. R. Strom. 1988. Molecular cloning, physical mapping and expression of *bet* genes governing the osmoregulatory choline-glycine betaine pathway of *Escherichia coli*. *J. Gen. Microbiol.* 134:1737-1746.
- Asahina, E., and K. Tanno. 1964. A large amount of trehalose in a frost-resistant insect. *Nature (London)* 204:1222.
- Attfield, P. V. 1987. Trehalose accumulates in *Saccharomyces cerevisiae* during exposure to agents that induce heat shock response. *FEBS Lett.* 225:259-263.
- Bachmann, B. J. 1990. Linkage map of *Escherichia coli* K-12, edition 8. *Microbiol. Rev.* 54:130-197.
- Bachmann, B. J. Personal communication.
- Becerra de Laredo, L., J. Ratouchniak, and F. Casse. 1977. Chromosomal location of gene governing the trehalose utilization in *Escherichia coli* K12. *Mol. Gen. Genet.* 152:105-108.
- Bochner, B. R., H.-C. Huang, G. L. Schieven, and B. N. Ames. 1980. Positive selection for loss of tetracycline resistance. *J.*



- Bacteriol. 143:926-933.
8. Boos, W., U. Ehmann, E. Bremer, A. Middendorf, and P. Postma. 1987. Trehalase of *Escherichia coli*. Mapping and cloning of its structural gene and identification of the enzyme as a periplasmic protein induced under high osmolarity growth conditions. *J. Biol. Chem.* 262:13212-13218.
  9. Boos, W., U. Ehmann, H. Forkl, W. Klein, M. Rimmick, and P. Postma. 1990. Trehalose transport and metabolism in *Escherichia coli*. *J. Bacteriol.* 172:3450-3461.
  10. Botsford, J. L. 1990. Analysis of protein expression in response to osmotic stress in *Escherichia coli*. *FEMS Microbiol. Lett.* 72:355-360.
  11. Cairney, J., I. R. Booth, and C. F. Higgins. 1985. *Salmonella typhimurium*: *proP* gene encodes a transport system for the osmoprotectant betaine. *J. Bacteriol.* 164:1218-1223.
  12. Chang, A. C. Y., and S. N. Cohen. 1978. Construction and characterization of amplifiable multicopy DNA cloning vehicles derived from the P15A cryptic miniplasmid. *J. Bacteriol.* 134:1141-1156.
  13. Chung, C. T., S. L. Niemela, and R. H. Miller. 1989. One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. *Proc. Natl. Acad. Sci. USA* 86:2172-2175.
  14. Crowe, J. H., L. M. Crowe, and D. Chapman. 1984. Preservation of membranes in anhydrobiotic organisms: the role of trehalose. *Science* 223:701-703.
  15. Csoska, L. 1982. A third L-proline permease in *Salmonella typhimurium* which functions in media of elevated osmotic strength. *J. Bacteriol.* 151:1433-1443.
  16. Diaz-Guerra, I., F. Moreno, and J. L. San Millan. 1989. *appR* gene product activates transcription of microcin C<sup>r</sup> plasmid genes. *J. Bacteriol.* 171:2906-2908.
  17. Esboo, M. W. 1988. *lac* fusion analysis of the *bet* genes of *Escherichia coli*: regulation by osmolarity, temperature, oxygen, choline, and glycine betaine. *J. Bacteriol.* 170:5208-5215.
  18. Galinski, E. A., and R. M. Herzog. 1990. The role of trehalose as a substitute for nitrogen-containing compatible solutes (*Ecithiorhodospira halochloris*). *Arch. Microbiol.* 153:607-613.
  19. Glaever, H. M., O. B. Styrvold, I. Kaasen, and A. R. Strom. 1988. Biochemical and genetic characterization of osmoregulatory trehalose synthesis in *Escherichia coli*. *J. Bacteriol.* 170:2841-2849.
  20. Gutierrez, C., M. Ardourel, E. Bremer, A. Middendorf, W. Boos, and U. Ehmann. 1989. Analysis and DNA sequence of the osmoregulated *treA* gene encoding the periplasmic trehalase of *Escherichia coli* K12. *Mol. Gen. Genet.* 217:347-354.
  21. Herbert, D., P. J. Phipps, and R. E. Strange. 1971. Chemical analysis of microbial cells, p. 209-344. In J. R. Norris and D. W. Ribbons (ed.), *Methods in microbiology*, vol. 5B. Academic Press, Inc. (London), Ltd., London.
  22. Hino, A., K. Mihara, K. Nakashima, and H. Takama. 1990. Trehalose levels and survival ratio of freeze-tolerant versus freeze-sensitive yeasts. *Appl. Environ. Microbiol.* 56:1386-1391.
  23. Hoeltzle, I., and J. G. Streeter. 1990. Increased accumulation of trehalose in rhizobia cultured under 1% oxygen. *Appl. Environ. Microbiol.* 56:3213-3215.
  24. Hottiger, T., T. Boller, and A. Wiemken. 1987. Rapid changes of heat and desiccation tolerance correlated with changes of trehalose content in *Saccharomyces cerevisiae* cells subjected to temperature shifts. *FEBS Lett.* 220:113-115.
  25. Jenkins, D. E., S. A. Chaisson, and A. Matin. 1990. Starvation-induced cross protection against osmotic challenge in *Escherichia coli*. *J. Bacteriol.* 172:2779-2781.
  26. Jovanovich, S. B., M. T. Record, Jr., and R. R. Burgess. 1989. In an *Escherichia coli* coupled transcription-translation system, expression of the osmoregulated gene *proU* is stimulated at elevated potassium concentration and by an extract from cells grown at high osmolarity. *J. Biol. Chem.* 264:7821-7825.
  27. Klein, W., U. Ehmann, and W. Boos. 1991. The repression of trehalose transport and metabolism in *Escherichia coli* by high osmolarity is mediated by trehalose-6-phosphate phosphatase. *Res. Microbiol.* 142:359-371.
  28. Kohara, Y., K. Akiyama, and K. Isono. 1987. The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. *Cell* 50:495-508.
  29. Laere, A. V. 1989. Trehalose, reserve and/or stress metabolite? *FEMS Microbiol. Rev.* 63:201-210.
  30. Lange, R., and R. Hengge-Aronis. 1991. Identification of a central regulator of stationary-phase gene expression in *Escherichia coli*. *Mol. Microbiol.* 5:49-59.
  31. Larsen, P. I., L. K. Sydnos, B. Landfald, and A. R. Strom. 1987. Osmoregulation in *Escherichia coli* by accumulation of organic osmolytes: betaines, glutamic acid, and trehalose. *Arch. Microbiol.* 147:1-7.
  32. Loewen, P. C., and B. L. Triggs. 1984. Genetic mapping of *kaiF*, a locus that with *kaiE* affects the synthesis of a second catalase species in *Escherichia coli*. *J. Bacteriol.* 160:668-675.
  33. Londesborough, J., and O. Vuorio. 1991. Trehalose-6-phosphate synthase/phosphatase complex from bakers' yeast: purification of a proteolytically activated form. *J. Gen. Microbiol.* 137:323-330.
  34. Mackenzie, K. F., K. K. Singh, and A. D. Brown. 1988. Water stress plating hypersensitivity of yeasts: protective role of trehalose in *Saccharomyces cerevisiae*. *J. Gen. Microbiol.* 134:1661-1666.
  35. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
  36. Martin, M. C., L. A. Diaz, M. B. Manzana, and C. Hardisson. 1986. Role of trehalose in the spores of *Streptomyces*. *FEMS Microbiol. Lett.* 35:49-54.
  37. Matin, A. 1991. The molecular basis of carbon-starvation-induced general resistance in *Escherichia coli*. *Mol. Microbiol.* 5:3-10.
  38. Miller, J. H. 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
  39. Mulvey, M. R., and P. C. Loewen. 1989. Nucleotide sequence of *kaiF* of *Escherichia coli* suggests KaiF protein is a novel  $\sigma$  transcription factor. *Nucleic Acids Res.* 17:9979-9991.
  40. Mulvey, M. R., J. Switala, A. Borys, and P. C. Loewen. 1990. Regulation of transcription of *kaiE* and *kaiF* in *Escherichia coli*. *J. Bacteriol.* 172:6713-6720.
  41. Piper, P. W., and A. Lockheart. 1988. A temperature-sensitive mutant of *Saccharomyces cerevisiae* defective in the specific phosphatase of trehalose biosynthesis. *FEMS Microbiol. Lett.* 49:245-250.
  42. Pradet, E., and P. L. Boquet. 1989. Mapping of the *Escherichia coli* acid glucose-1-phosphatase gene *agg* and analysis of its expression in vivo by use of an *agg-phoA* protein fusion. *J. Bacteriol.* 171:3511-3517.
  43. Prince, W. S., and M. R. Villarejo. 1990. Osmotic control of *proU* transcription is mediated through direct action of potassium glutamate on the transcription complex. *J. Biol. Chem.* 265:17673-17679.
  44. Ramirez, R. M., and M. Villarejo. 1991. Osmotic signal transduction to *proU* is independent of DNA supercoiling in *Escherichia coli*. *J. Bacteriol.* 173:879-885.
  45. Reed, R. H., L. J. Borowitzka, M. A. Mackay, J. A. Chudek, R. Foster, S. R. C. Warr, D. J. Moore, and W. D. P. Stewart. 1986. Organic solute accumulation in osmotically stressed cyanobacteria. *FEMS Microbiol. Rev.* 39:51-56.
  46. Roberts, I., R. Mountford, N. High, D. Bitter-Suermann, K. Jann, K. Timmis, and G. Boulnois. 1986. Molecular cloning and analysis of genes for production of K5, K7, K12, and K92 capsular polysaccharides in *Escherichia coli*. *J. Bacteriol.* 168:1228-1233.
  47. Rod, M. L., K. Y. Alam, P. R. Cunningham, and D. P. Clark. 1988. Accumulation of trehalose by *Escherichia coli* K-12 at high osmotic pressure depends on the presence of amber suppressors. *J. Bacteriol.* 170:3601-3610.
  48. Schweizer, H., and W. Boos. 1983. Transfer of the  $\Delta(\argF-lac)U169$  mutation between *Escherichia coli* strains by selection for a closely linked Tn10 insertion. *Mol. Gen. Genet.* 192:293-294.

49. Silhavy, T. J., M. L. Berman, and L. W. Enquist. 1984. Experiments with gene fusions. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
50. Strøm, A. R., P. Falkenberg, and B. Landfald. 1986. Genetics of osmoregulation in *Escherichia coli*: uptake and biosynthesis of organic osmolytes. FEMS Microbiol. Rev. 39:79-86.
51. Stüber, D., and H. Bujard. 1981. Organization of transcriptional signals in plasmids pBR322 and pACYC184. Proc. Natl. Acad. Sci. USA 78:167-171.
52. Styrvoid, O. B., and A. R. Strøm. 1991. Synthesis, accumulation, and excretion of trehalose in osmotically stressed *Escherichia coli* K-12 strains: influence of amber suppressors and function of the periplasmic trehalase. J. Bacteriol. 173:1187-1192.
53. Touati, E., E. Dassa, and P. L. Boquet. 1986. Pleiotropic mutations in *appR* reduce pH 2.5 acid phosphatase expression and restore succinate utilisation in CRP-deficient strains of *Escherichia coli*. Mol. Gen. Genet. 202:257-264.
54. Touati, E., E. Dassa, J. Dassa, P. L. Boquet, and D. Touati. 1991. Are *appR* and *kutF* the same *Escherichia coli* gene encoding a new sigma transcription initiation factor? Res. Microbiol. 142: 29-36.

# A novel methyl transferase induced by osmotic stress in the facultative halophyte *Mesembryanthemum crystallinum*

Claim 66  
Myoinositol O-methyltransferase

Daniel M. Vernon<sup>1</sup> and Hans J. Bohnert<sup>1,2</sup>

Departments of <sup>1</sup>Molecular and Cellular Biology and <sup>2</sup>Biochemistry and of Plant Sciences, University of Arizona, Tucson, AZ 85721, USA

Communicated by C.J. Leaver

Molecular mechanisms of osmotic stress tolerance were studied in *Mesembryanthemum crystallinum* (ice plant), a facultative halophyte capable of adjusting to and surviving in highly saline conditions. We screened a subtracted cDNA library enriched for salt stress-induced mRNAs to identify transcripts involved in this plant's adaptation to salinity. One mRNA, *Imt1*, was found to be up-regulated in leaves and, transiently, in roots. Nuclear run-on assays indicated that this mRNA is transcriptionally regulated. *Imt1* encoded a predicted polypeptide of M<sub>r</sub> 40 250 which exhibited sequence similarity to several hydroxymethyl transferases. Expression of the protein in *Escherichia coli* and subsequent activity assays identified the protein as a novel myo-inositol O-methyl transferase which catalyzes the first step in the biosynthesis of the cyclic sugar alcohol pinitol. Pinitol accumulates in salt-stressed *M. crystallinum* and is abundant in a number of salt- and drought-tolerant plants. The presence of high levels of sugar alcohols correlates with osmotolerance in a diverse range of organisms, including bacteria, fungi and algae, as well as higher plants. The stress-initiated transcriptional induction of IMT1 expression in a facultative halophyte provides strong support for the importance of sugar alcohols in establishing tolerance to osmotic stress in higher plants.

**Key words:** gene expression/myo-inositol O-methyl transferase/osmoprotection/salt stress/sugar alcohol

## Introduction

For plants, environmental stress is an inescapable and persistent condition. Water stress imparted by drought or high salinity is the most widespread abiotic stress and constitutes the most stringent factor in limiting plant distribution and productivity (Boyer, 1982). A large body of physiological work has described how various species respond to water deficit. A number of salt tolerant plants (halophytes) have evolved strategies that enable them to adjust to salinity, avoid salinity-induced water stress and succeed in environments that will not accommodate salt-sensitive species (glycophytes). Such strategies can include changes in photosynthesis, alterations in membrane structure, exclusion of salts, accumulation and partitioning of ions, increases in intracellular organic solutes and precocious entry into reproductive phase and senescence (Jefferies, 1981).

There have been a number of recent investigations into the molecular responses of plants to water stress. Most of these studies have identified polypeptides that are induced in glycophytic plants during water stress imposed by drought (Gomez *et al.*, 1988; Mundy and Chua, 1988; Mason and Mullet, 1990; Cohen *et al.*, 1991; Plant *et al.*, 1991). Bartels *et al.* (1990, 1991) have identified several mRNAs, including one encoding an aldose reductase, induced during drought in the resurrection plant, *Craterostigma plantagineum*, which can remain dormant yet viable during long periods of complete desiccation. A consistent theme in glycophytic water-stress responses is the expression of polypeptides which are also associated with abscisic acid (ABA)-mediated seed and embryo desiccation (Skriver and Mundy, 1990). Examples include Rab proteins (Mundy and Chua, 1988), Lea proteins (Dure *et al.*, 1989) and the aldose reductase mentioned above, which is induced in barley embryos as well as desiccating resurrection plant leaf tissue (Bartels *et al.*, 1991). The function of many of these proteins (e.g. Rabs and Leas) during stress is not clear. Many share some biochemical characteristics such as high hydrophilicity, and it is thought that their role in maturing seeds and stressed leaf tissue is to preserve the structural integrity of cells during the desiccation process (Skriver and Mundy, 1990).

The effects of salinity on glycophytes have also been investigated. Claes *et al.* (1990) identified a salinity-induced rice mRNA encoding a protein of unknown function. Work with tobacco in tissue culture has established that salt-adapted cell lines exhibit changes in ion compartmentation (Binzel *et al.*, 1988), cell wall chemistry (Iraki *et al.*, 1989) and gene expression (Singh *et al.*, 1989).

Studies with glycophytes have established the importance of gene expression in plant stress responses and defined classes of proteins that are induced by desiccation, but most of these investigations have not been successful in distinguishing pathological from stress-ameliorating responses. We have taken advantage of a unique system, the inducible halophyte *Mesembryanthemum crystallinum* (common ice plant), in order to understand the mechanisms by which some plants avoid salinity-induced water deficit. As a facultative halophyte, the ice plant undergoes a set of stress-inducible physiological and biochemical changes that allow it to adjust and maintain cell viability and turgor, conserve water and continue growth during extreme salt stress. One physiological adaptation, the stress-induced switch from C<sub>3</sub> to Crassulacean acid metabolism (CAM), has been studied in depth (Winter and von Willert, 1972; Bohnert *et al.*, 1988 and references therein). CAM is an alternative photosynthesis pathway that allows night-time carbon fixation, thereby resulting in decreased evaporative water loss (Ting, 1985). The shift to CAM has been shown to require increased expression of mRNAs encoding a number of CAM pathway enzymes (Ostrem *et al.*, 1987; Vernon *et al.*, 1988; Schmitt *et al.*, 1988; Ostrem *et al.*, 1990), establishing *Mesem-*

*bryanthemum* as an attractive system for the study of salt stress at the molecular level.

CAM is considered to be separate from the ice plant's ability to adjust to and tolerate high salinity at the cellular level. CAM is a complex physiological phenomenon that serves as a long-term water conservation strategy, and is not fully operative until 7–14 days following environmental stress in 6 week-old *M. crystallinum* (Winter and Gademann, 1991; Bohnert *et al.*, 1988). CAM is preceded by a period of transient wilting (2–4 days) followed by recovery and continued growth (Winter and Gademann, 1991; Bohnert *et al.*, 1988; D.Vernon, unpublished). It is likely that there are mechanisms distinct from CAM that are responsible for this observed adjustment to saline conditions. To investigate the molecular basis of salt tolerance, we generated and screened a subtracted cDNA library enriched for stress-induced sequences. Here we show that this facultative halophyte responds to osmotic stress by transcriptional induction of a gene encoding a novel methyl transferase. This methyl transferase is identified by functional assay as a myo-inositol *O*-methyl transferase involved in the biosynthesis of a cyclic sugar alcohol, pinitol. Pinitol is abundant in a number of salt- and drought-tolerant plant species, including salt-stressed *M. crystallinum*, where it can accumulate to >70% of soluble carbohydrate (Paul and Cockburn, 1989). The transcriptional induction of this cyclitol biosynthesis enzyme in salt-stressed *M. crystallinum* indicates that the cyclic sugar alcohol pinitol may play a crucial role during adaptation to osmotic stress in this facultative halophyte.

## Results

### cDNA isolation

To identify environmentally-induced molecular changes involved in the ice plant's adaptation to salinity, we constructed and differentially screened a subtracted cDNA library enriched for stress-induced sequences. cDNA was generated from poly(A)<sup>+</sup> RNA that had been isolated from 7 week-old soil-grown plants stressed with 500 mM NaCl for 10 h. By this time, plants start to recover from stress-induced transient wilting (Winter and Gademann, 1991), but have not yet accumulated large amounts of mRNAs encoding components of the well characterized CAM pathway (Ostrem *et al.*, 1987, 1990). Three cycles of differential screening of ~10<sup>5</sup> plaques with labeled first-strand cDNAs from stressed and unstressed plants yielded eight consistently up-regulated inserts. Cross-hybridization experiments indicated that the inserts represented three distinct clones (data not shown). A 1.6 kb insert, now referred to as *lmt1*, was chosen for further analysis.

### Expression of *lmt1* mRNA

Expression of the *lmt1* transcript was analyzed on Northern blots of RNA isolated from root and leaf tissue of hydroponically-grown plants (Figure 1). Total RNA was isolated from unstressed plants and plants harvested at several timepoints over the course of a 6 day stress with 400 mM NaCl. The *lmt1* cDNA probe hybridized to a salinity-induced mRNA of ~1.6 kb in both leaf and root RNA. The pattern of induction differed in these two organs. In unstressed leaves, *lmt1* was present at very low levels. The transcript accumulated gradually in stressed leaves (Figure 1A), being detectable after 6 h of stress, but inconspicuous until the

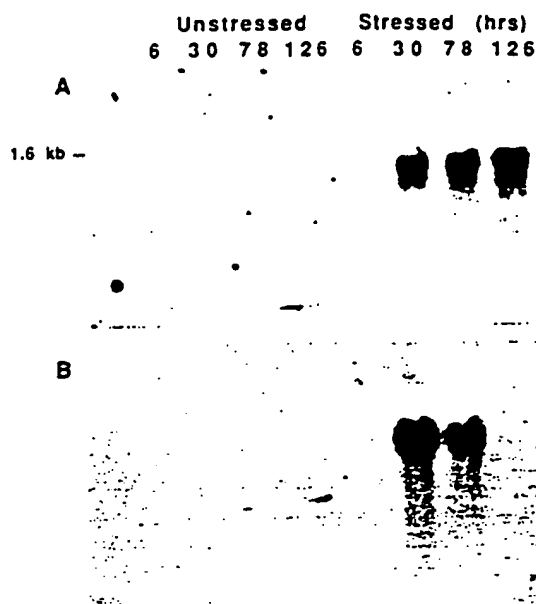


Fig. 1. Hybridization of the *lmt1* cDNA to Northern blots of *M. crystallinum* RNA. Total RNA was isolated from leaves (A) or roots (B) of hydroponically-grown 6 week-old control plants (unstressed) and plants that had been stressed with 400 mM NaCl for 6, 30, 78 and 126 h. Control plants were harvested at each timepoint. 10 µg of leaf (A) or 20 µg of root (B) RNA from each timepoint were resolved on formaldehyde-agarose gels, transferred to nitrocellulose and probed with <sup>32</sup>P-labeled *lmt1* cDNA. Autoradiographs of blots are shown. The root blot was over-exposed to illustrate the disappearance of the transcript in roots between 78 and 126 h of stress. Transcript size is indicated in kb.

second day, after 30 h of stress. Accumulation of the transcript in leaves reached a maximum by the sixth day of salt treatment (126 h). *lmt1* was transiently up-regulated in roots (Figure 1B), rising from undetectable levels to a maximum level of expression during the second day of stress (30 h). Interestingly, the mRNA completely disappeared from roots by the time-of maximum expression in leaves. Blots of dilutions of leaf and root RNA indicated that, at the times of maximal expression, the *lmt1* mRNA is >25 times more abundant in leaves than in roots (data not shown). A prolonged exposure of the root RNA Northern blot is shown in Figure 1 to illustrate the complete disappearance of this transcript between the fourth and sixth days of stress.

### Genomic Southern analysis

The *lmt1* transcript induced in leaves and roots is encoded by a single nuclear gene or small gene family. Nuclear genomic DNA digested with various restriction enzymes was resolved by 1% agarose gel and Southern-blotted along with genome copy number equivalents of the cloned cDNA (Figure 2). Blots were probed with <sup>32</sup>P-labeled cDNA fragments and signal intensities were quantified using a β-scanner (Betagen Inc., Waltham, MA). Probes specific to either the 5'-coding region (not shown) or the 3'-non-coding end of the cDNA hybridized to single bands of equal intensity in each lane (Figure 2A). High stringency wash conditions were identical to those used for the Northern blots shown in Figure 1. Comparison of band intensities with copy number reconstitutions (Figure 2B) indicated that the bands probably represent a single gene. A blot probed with a 400 bp 3'-end fragment is shown.

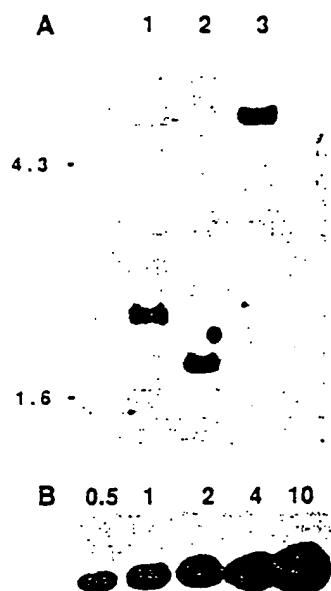


Fig. 2. Southern analysis of *M. crystallinum* genomic DNA. (A) 4  $\mu$ g of nuclear DNA were digested with *Eco*RI (lane 1), *Hind*III (lane 2) or *Hinc*II (lane 3), resolved by agarose gel electrophoresis and Southern blotted alongside known amounts of *lmt1* cDNA corresponding to 0.5, 1, 2, 4 and 10 copies per genome (B). The blot shown was probed with a  $^{32}$ P-labeled 400 bp 3'-end fragment of the *lmt1* cDNA. Positions of size markers are shown in kb.

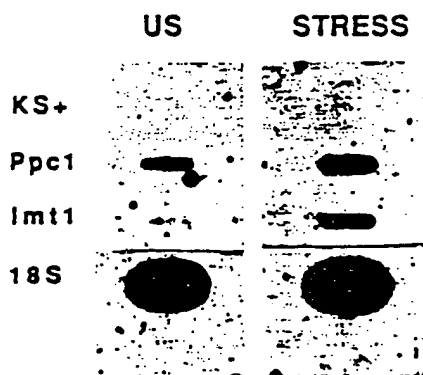


Fig. 3. Nuclear run-on assays of *lmt1* transcription. *lmt1* cDNA was slot-blotted onto nitrocellulose and probed with  $^{32}$ P-labeled transcripts generated by nuclei isolated from unstressed and salt-stressed (102 h) *M. crystallinum*. A PEP carboxylase cDNA clone (*Ppc1*) was included on blots as a positive control for transcriptional induction (Cushman *et al.*, 1989). Bluescript DNA (KS<sup>+</sup>) served as a control for background. Hybridization to 18S rDNA (18S) was used to normalize total transcription rates of nuclei from stressed and unstressed plants.

#### Nuclear run-on assays

The *lmt1* mRNA is transcriptionally induced in leaf tissue (Figure 3). Nuclear run-on experiments were performed with nuclei isolated from control and salt-stressed *M. crystallinum*. Transcripts were labeled with [ $^{32}$ P]UTP and hybridized to excess *lmt1* cDNA slot-blotted onto nitrocellulose. *Ppc1* DNA, which encodes the transcriptionally induced CAM enzyme PEP carboxylase (Cushman *et al.*, 1989), was included as a positive control. Hybridization to 18S rDNA was used to normalize the overall transcription rates of

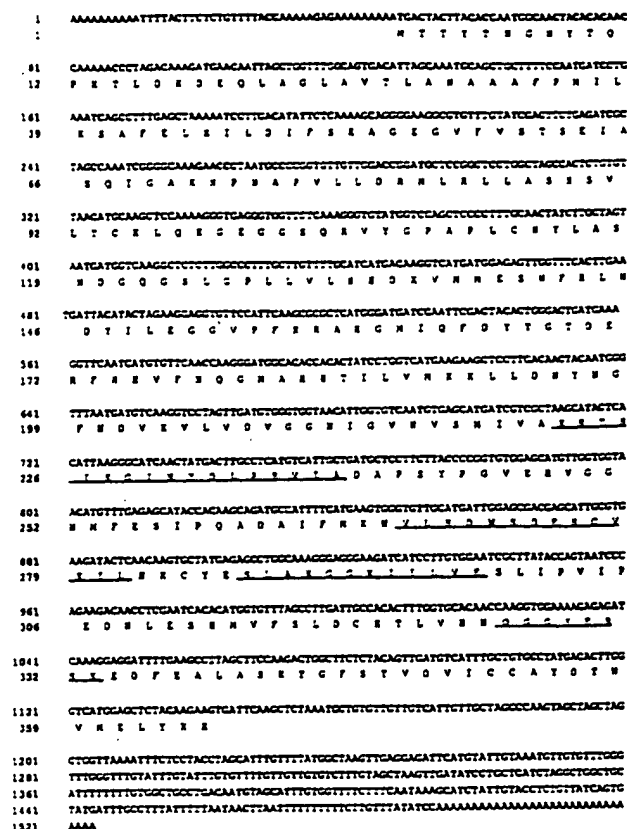


Fig. 4. Nucleotide sequence and predicted amino acid sequence of the *lmt1* cDNA. Underlined regions share homology with hydroxymethyl transferases from plant, bacterial and mammalian sources (Bugos *et al.*, 1991). The sequence data of *lmt1* have been deposited in the DDBJ/GenBank Nucleotide Sequence Databases under the accession number M87340.

different reactions.  $\beta$ -scan analysis of experiments such as that shown in Figure 3 indicated that transcription of the *lmt1* message is four to five times higher in leaves from salt-stressed plants than in leaves of unstressed control plants of the same age. This increase is comparable to, and perhaps greater than, the stress-induced increase in transcription rate observed for the *Ppc1* transcript (Figure 3, Cushman *et al.*, 1989).

#### Sequence analysis

To gain insight into the possible biochemical and physiological functions of the IMT1 protein, the sequence of the cDNA was determined (Figure 4). The clone was 1524 bp long. It contained an A+T-rich leader sequence and an ATG start codon followed by an uninterrupted reading frame of 1095 nucleotides. The 3' end consisted of a long non-coding region of 383 nucleotides, which included two possible polyadenylation recognition sequences (AATAA and AATAAA) located 28 and 74 bases, respectively, upstream of a 31 base poly(A) tail. Analysis of the *lmt1* sequence predicted a hydrophilic polypeptide of 365 amino acids with a molecular mass of 40 250 (Figure 4). A search of the NBRF database revealed similarity to a bovine pineal gland hydroxyindole O-methyltransferase (Ishida *et al.*, 1987). Homology (including conservative amino acid replacements) was 55% over the entire length. Significant identity (27%) was confined to 195 amino acids in the C-terminal portion of the protein (alignments not shown). The predicted IMT1

polypeptide was found to be even more closely related to two plant bifunctional hydroxymethyl transferases which methylate the lignin monomers caffeic acid and hydroxyferulic acid (Bugos *et al.*, 1991; Gowri *et al.*, 1991). Comparison with these proteins revealed >50% identity over the entire length of the IMT1 sequence (not shown). Regions of shared homology with *S*-adenosyl-L-methionine (SAM)-dependent hydroxymethyl transferases (OMTs) are underlined in Figure 4 (Bugos *et al.*, 1991).

#### Expression of IMT1 in *Escherichia coli* and identification of activity

Although sequence alignments suggested a general biochemical function for the IMT1 protein, amino acid identities were not sufficient for the identification of the enzyme's substrate specificity. We hypothesized that one possible role for a methyl transferase in the salt stress response in *M. crystallinum* was in the biosynthesis of pinitol (1D-3-*O*-methyl *chiro*-inositol). Pinitol is a methylated cyclic sugar alcohol (cyclitol) that accumulates to high levels in salt-stressed *M. crystallinum* (Paul and Cockburn, 1989). Its appearance and accumulation parallel that of the *Imt1* mRNA (Figure 1) during the initial days of salt stress (Paul and Cockburn, 1989). Pinitol has potentially osmoprotective properties and is thought to serve as an intracellular osmolyte (possibly in the chloroplasts) during osmotic adjustment in this inducible halophyte (see Discussion). Pinitol is derived from the six-carbon cyclic sugar alcohol myo-inositol (Figure 5). Its synthesis involves an initial methylation of myo-inositol by a SAM-dependent hydroxymethyl trans-

ferase, followed by an enzyme-catalyzed epimerization via a *keto* intermediate. In angiosperms, the pathway proceeds via ononitol (1D-4-*O*-methyl myo-inositol; Figure 5). An isomer of ononitol, sequoyitol (1D-5-*O*-methyl myo-inositol), is the methylated intermediate in gymnosperms (Dittrich and Brandl, 1987).

To substantiate the hypothesized physiological role of the IMT1 protein in pinitol biosynthesis, the polypeptide was expressed in *E. coli* and bacterial lysates were tested for myo-inositol hydroxymethyl transferase activity. As is evident from Figure 6A, *E. coli* containing the clone in the proper orientation expressed a protein of molecular mass ~40 kDa that was not present in control cells (containing the clone in the opposite orientation) that were grown under identical conditions. This protein co-migrated on SDS-polyacrylamide gels with the reticulocyte system *in vitro* translation product of the *in vitro*-transcribed *Imt1* clone (Figure 6B). Expression ranged from ~1% to >10% of *E. coli* protein, depending on growth conditions.

We tested *E. coli* extracts from control cells and cells expressing the IMT1 protein for myo-inositol-dependent *O*-methyl transferase activity in an assay using  $^{14}\text{CH}_3$ -labeled SAM as a methyl group donor. Carbohydrates from post-assay extracts were separated by HPLC and radioactive fractions identified by scintillation counting. A radioactive product with a retention time of ~11.1 min was generated by extracts from cells expressing the IMT1 protein, but not by control extracts (Figure 7). The appearance of this peak was dependent on the addition of myo-inositol to the assays.

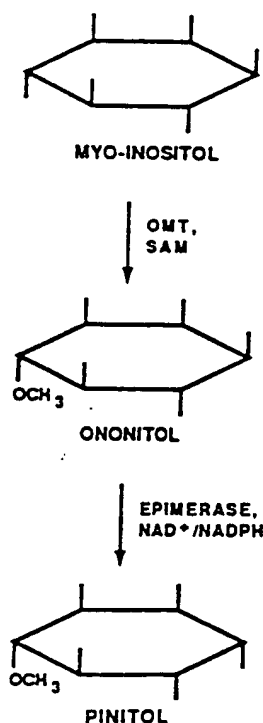


Fig. 5. Pathway of pinitol biosynthesis in angiosperms. Pinitol (1D-3-*O*-methyl *chiro*-inositol) is synthesized from myo-inositol, a six-carbon ring polyol derived from glucose-6-phosphate. The synthesis pathway consists of a methylation step catalyzed by a position-specific *S*-adenosyl-L-methionine (SAM)-dependent *O*-methyl transferase (OMT), followed by an enzyme-catalyzed epimerization which may require NAD<sup>+</sup> and/or NADPH (Loewus and Dickinson, 1982). The methylated intermediate is ononitol. Vertical bars on ring structures represent hydroxyl groups.

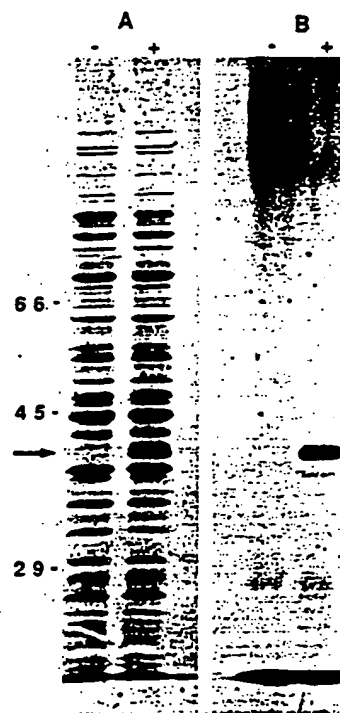


Fig. 6. Expression of IMT1 in *E. coli*. The *Imt1* cDNA expression constructs were transformed into *E. coli* BL21(DE3) cells as described in Materials and methods. (A) SDS-PAGE of proteins from *E. coli* transformed with *Imt1* cloned 3'-5' (-) or 5'-3' (+) behind the bluescript T7 promoter. (B) SDS-PAGE of *in vitro* translation products of unprogrammed reticulocyte lysate translation system (-) or lysate programmed with *in vitro*-transcribed *Imt1* RNA (+). All samples in panels A and B were run on the same gel, which was cut and either Coomassie stained (A) or fluorographed (B). Positions of size markers are labeled in kDa. Arrow indicates position of IMT1.

Two less prominent peaks were generated by all assays (regardless of myo-inositol addition or IMT1 expression), suggesting the presence of endogenous SAM-dependent OMTs in *E. coli*. The activity was sensitive to both heat (70°C, 15 min) and proteinase K treatment, and  $^{14}\text{C}$

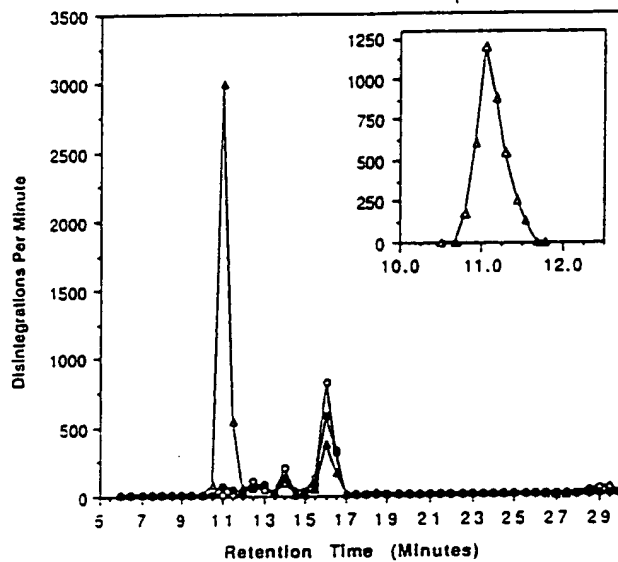


Fig. 7. Myo-inositol *O*-methyl transferase assays of *E. coli* extracts. Equal amounts of total soluble protein from cells expressing IMT1 were assayed with (Δ) or without (●) myo-inositol. Control extracts from cells transformed with only Bluescript (○) were assayed with myo-inositol. [ $^{14}\text{C}$ ]methyl-SAM was included as a methyl group donor. A portion (8%) of post-assay carbohydrates were resolved by HPLC as described in Materials and methods.  $^{14}\text{C}$  d.p.m. of 0.5 min fractions from representative HPLC runs are shown. Inset: to define more precisely the retention time of the IMT1 product, 7.5 s fractions with retention times between 10.5 and 12.0 min were collected and analyzed.

incorporation into the product was linear during the course of the assay (data not shown). These assays indicated that *Imt1* encodes a SAM-dependent myo-inositol *O*-methyl transferase.

The  $^{14}\text{C}$ -labeled product of the assays on *E. coli* extracts was visible on HPLC traces as a distinct peak with a retention time of slightly under 11.1 min (Figure 8, panel 3) that was not present in assays from control extracts (Figure 8, panels 1 and 2). To establish that the methylated myo-inositol generated by the IMT1 protein was ononitol, the methylated intermediate in pinitol biosynthesis, its retention time was compared with that of methyl-myoinositol standards. There are four methyl-myoinositol isomers: sequoyitol, ononitol and *D*- and *L*-bornesitol (1*D*- or 1*L*-*O*-methyl myo-inositol) (Loewus and Loewus, 1980). Only ononitol and sequoyitol are possible precursors for pinitol (Dittrich and Brandl, 1987; Figure 5), and only ononitol has been documented as the precursor to pinitol in angiosperms (Dittrich and Brandl, 1987). Extracts from control *E. coli* (such as that shown in Figure 8, panel 1) were spiked with standards, processed as assays (see Materials and methods) and analyzed by HPLC. The retention times of sequoyitol and bornesitol were 11.5 and 12.2 min respectively (data not shown). Ononitol, however, displayed a retention time identical to that of the reaction product (Figure 8, panel 4), indicating that the myo-inositol-dependent *O*-methyl transferase encoded by the *Imt1* mRNA does indeed generate the methylated intermediate in the biosynthesis of pinitol.

#### IMT1 activity in plant extracts

To verify that the induction of the *Imt1* mRNA in stressed *M. crystallinum* is accompanied by an increase in IMT1 expression, extracts from leaves of salt stressed (102 h) and unstressed *M. crystallinum* were assayed for IMT1 enzyme activity (Figure 9). Carbohydrates from post-assay plant extracts were resolved by HPLC, fractions were collected

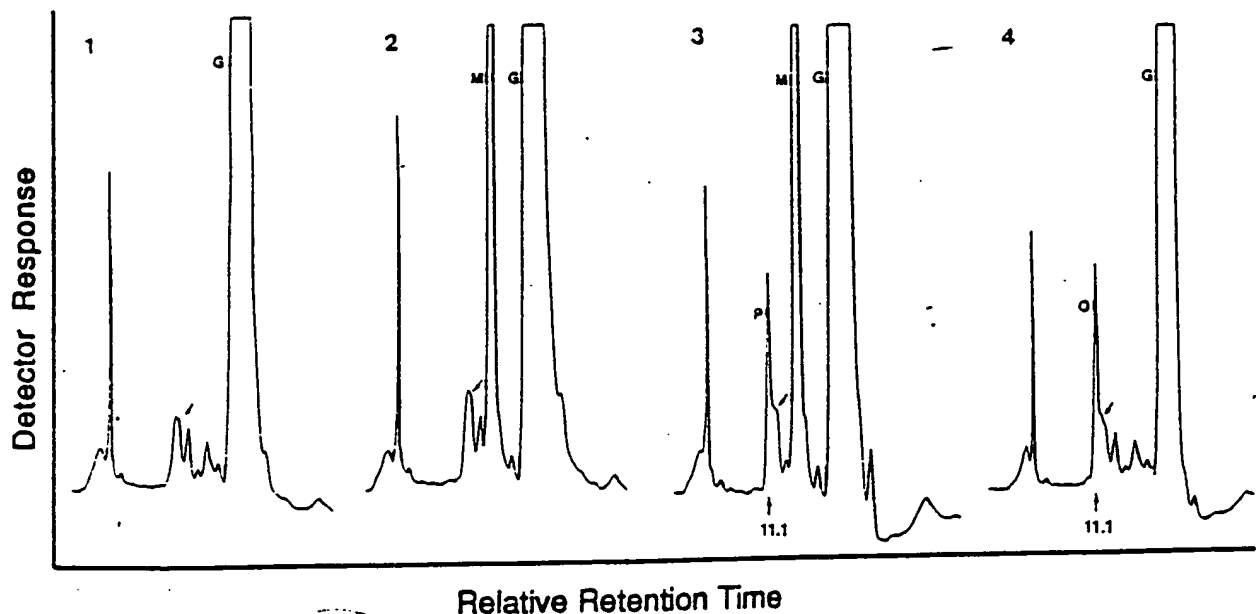


Fig. 8. Identification of the methyl-myoinositol product of IMT1. Representative HPLC traces of carbohydrates from post-assay *E. coli* extracts are shown: Panel 1: assay of extract containing IMT1 but no myo-inositol. Panel 2: assay (with myo-inositol) of extract from control cells transformed with vector only. Panel 3: assay of extract containing IMT1 protein plus myo-inositol. Panel 4: same extract as panel 1, but spiked with 1.5 nmol of ononitol standard. Peak labels: G, glycerol added to protein extracts for storage; M, myo-inositol added to extracts for assay; P, assay product; O, ononitol spike. Tilted arrows indicate the position of an endogenous *E. coli* peak present in all extracts, useful here as a marker. Retention times of product and of ononitol standard are indicated.

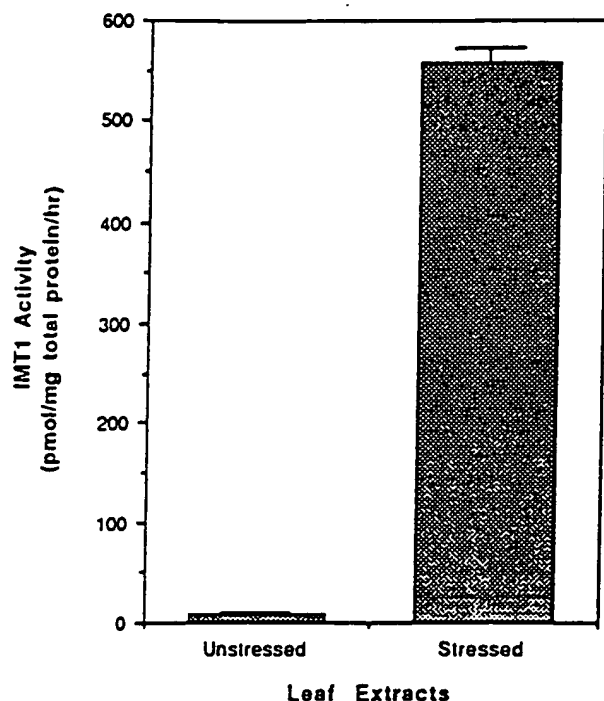


Fig. 9. IMT1 activity in stressed and unstressed *M. crystallinum*. Crude soluble protein extracts from leaves of unstressed plants (unstressed) or plants stressed for 102 h (stressed) were assayed for myo-inositol OMT activity in assays similar to those run on *E. coli* extracts (Figure 7). Assays were run for 2 h, during which time  $^{14}\text{C}$  incorporation into product was linear (data not shown). Proteins were extracted from leaf material pooled from three stressed or three unstressed plants. Activity values are the mean of three separate assays; standard deviations are shown. No activity was detected in assays carried out without myo-inositol (data not shown).

and c.p.m. determined by scintillation counting.  $^{14}\text{C}$  c.p.m. were detected in a single peak with a retention time of  $\sim 11.1$  min and the appearance of  $^{14}\text{C}$  in carbohydrate fractions was dependent on the addition of myo-inositol to assays (data not shown). Stressed plants exhibited a large ( $\sim 60$ -fold) increase in IMT1 activity over unstressed plants, which contained an almost undetectable level of activity (Figure 9).  $^{14}\text{C}$  incorporation into product was linear during the course of the assay (data not shown). These assays indicated that *Imt1* mRNA induction results in increased IMT1 activity in stressed-plant extracts, suggesting that the increased expression of this transcript is, at least in part, responsible for the concurrent accumulation of pinitol observed in salt-stressed *M. crystallinum*.

## Discussion

Halophytic plants have developed mechanisms of osmoregulation that allow them to avoid salinity-induced water stress. We have used a facultative halophyte, *M. crystallinum*, to investigate the molecular basis of salt tolerance. When subjected to salinity stress, *M. crystallinum* undergoes a brief period of wilting followed by recovery and continued growth (Bohnert *et al.*, 1988; Winter and Gademann, 1991). In an attempt to elucidate the biochemical mechanisms responsible for this observed adjustment to stress, we generated a cDNA library enriched for transcripts induced during the initial hours of stress. Differential screening of this library yielded a dramatically up-regulated mRNA, *Imt1*, encoding a polypeptide with homology to several methyl transferases.

Expression of the protein in *E. coli* and activity assays identified the enzyme as a myo-inositol *O*-methyl transferase. This enzyme methylates myo-inositol to produce ononitol, which is subsequently epimerized by a different enzyme to produce pinitol, a cyclic sugar alcohol known to accumulate to high levels in salt-stressed *M. crystallinum* (Paul and Cockburn, 1989). Enzyme assays performed on plant extracts established that the induction of *Imt1* mRNA is accompanied by the appearance of myo-inositol *O*-methyl transferase activity in salt-stressed plants. To our knowledge, this study is the first example of differential screening of a cDNA library leading to the biochemical identification of a novel plant enzyme with a defined physiological function.

Pinitol is one of a larger class of compounds, sugar alcohols, whose presence correlates with osmotic stress tolerance in a wide range of organisms. Cyclic sugar alcohols such as pinitol and straight-chain polyols, such as mannitol and sorbitol, are found in high levels in desiccation- and salt-tolerant bacteria, fungi and algae as well as in halophytic higher plants (Csonka and Hanson, 1991; Bielecki, 1982; Hellebust, 1976). Pinitol is the most common of the various myo-inositol-derived cyclitols found in plants (Dittrich and Korak, 1984). It is found at high levels in a diverse set of salt-tolerant plant species, including maritime pine (Nguyen and Lamant, 1988), several legumes (Gorham *et al.*, 1988), halophytic mangroves (Popp, 1984) and both dicots and monocots from maritime habitats (Gorham *et al.*, 1981). In *M. crystallinum*, pinitol can accumulate to  $> 70\%$  of total soluble carbohydrate (up to 10% of plant dry weight) following NaCl stress (Paul and Cockburn, 1989). It may reach intracellular concentrations of up to 300 mM. Localization studies by Paul and Cockburn (1989) suggest that  $> 60\%$  of the pinitol in stressed *M. crystallinum* is in the cytoplasm, with the rest residing in the chloroplasts.

The exact physiological role of sugar alcohols during osmotic stress has not been determined. One view is that sugar alcohols are involved in osmoregulation: they accumulate to high levels to act as intracellular osmolytes that balance external osmotic pressure. In many osmotolerant fungi and algae, sugar alcohol levels are variable and correlate with the osmotic potential of the extracellular environment (Bielecki, 1982; Hellebust, 1976). In halophytic higher plants, sugar alcohols are thought to accumulate in the cytoplasm (and/or chloroplasts) to counteract the high osmotic potential of the vacuoles, which selectively accumulate salt to maintain turgor and sequester ions away from the cytosol, where they would be harmful to enzyme activities (McCue and Hanson, 1990; Pollard and Wyn Jones, 1979; Hellebust, 1976). A similar role has been proposed for other low molecular weight metabolites such as proline and glycine-betaine, which accumulate in many environmentally-stressed plants (McCue and Hanson, 1990; Pollard and Wyn Jones, 1979). Compounds that accumulate to high levels to serve as osmolytes in the cytoplasm must be compatible with cellular function. They must be non-disruptive to enzyme structure and activity even when present in high concentration, and they must be non-reactive (Yancey *et al.*, 1982; Gorham *et al.*, 1981; Hellebust, 1976). Also, osmolytes that accumulate to high intracellular concentration must be removed from the flux through major metabolic pathways. Metabolically inert, non-reactive cyclitols such as pinitol fit all of these criteria.

Osmoregulation is only one of the roles suggested for sugar



alcohols in stress-tolerant organisms. It has also been proposed that their primary purpose is one of macromolecular osmoprotection (Schobert, 1977). *In vitro* studies indicate that these compounds may have hydrogen bonding characteristics that allow them to protect macromolecules from the adverse effects of excessive water loss, high temperature and increased ionic strength in the surrounding medium (Laurie and Stewart, 1990; Bielecki, 1982; Ahmad *et al.*, 1979), presumably by tightly associating with proteins and membrane components to replace water lost during desiccation (Yancey *et al.*, 1982; Schobert, 1977). For this purpose, sugar alcohols are thought to be important in desiccation-tolerant organisms such as yeast (Coutinho *et al.*, 1988) and, possibly, the resurrection plant *Craterostigma plantagineum* (Bartels *et al.*, 1991). It is likely that sugar alcohols can carry out either osmoregulatory or osmoprotective roles in the various organisms where they are found, depending on the level to which they accumulate, their intracellular location and the environmental stress being encountered.

Although polyol accumulation has been correlated with osmotolerance in a number of plants, the mechanisms of accumulation have not been clear (Loewus and Dickinson, 1982). It has been noted that the higher levels of cyclitols and other putative intracellular osmolytes seen in stressed or salt-tolerant plants could, rather than being a beneficial response, be a consequence of 'impaired metabolism' during stress (Wyn Jones and Gorham, 1983). The transcriptional induction of the myo-inositol OMT (Figure 3) during adjustment to saline conditions indicates that sugar alcohol accumulation in the ice plant is not an artifact of disturbed metabolism. Rather, it appears to be part of a genetically determined, adaptive response that enables *M. crystallinum* to adjust physiologically to extreme osmotic stress.

The IMT enzyme is one of four distinct *O*-methyl transferases that methylate myo-inositol at specific hydroxyl positions to create ononitol, sequoyitol, or L- or D-bornesitol (Loewus and Loewus, 1980; Hoffmann-Ostenhoff *et al.*, 1978). Although two of these enzymes (those that methylate myo-inositol to make L- and D-bornesitol) have been purified (Hoffmann-Ostenhoff, 1978), none have been extensively biochemically characterized and no myo-inositol OMT sequences have been determined prior to this study. That these enzymes are highly specific for their site of methylation is supported by the work of Hoffmann-Ostenhoff (1978), as well as by our assay results, which generated only ononitol (Figures 7 and 8; and data not shown). Cloning and sequence analysis of other myo-inositol OMTs, as well as more in-depth biochemical analysis of the ice plant OMT described here, would provide information on the comparative biochemistry and specificity of these enzymes. The over-expression of IMT1 in *E. coli* (Figure 6) is a first step for both of these endeavors.

The transient expression of *lmt1* in roots (Figure 1) provided the first indication that the *lmt1* transcript was involved in a non-organ-specific stress response distinct from the leaf-specific switch to CAM photosynthesis that occurs in salt-stressed *M. crystallinum*. However, the different patterns of *lmt1* expression in leaves and roots (Figure 1) and the much greater abundance of the transcript in leaf tissue may indicate that the cells in these organs have distinct mechanisms of osmoregulation or that the roots play a different role in the response to stress at the whole-plant

level. Indeed, it has been noted in physiological studies that the root cortical cells of *M. crystallinum* do not fully regain turgor following NaCl stress, while leaves recover turgor after 4 days of stress (Winter and Gademann, 1991). The greater accumulation of *lmt1* mRNA in leaves reflects what has previously been observed for pinitol levels in the ice plant (Paul and Cockburn, 1989). Higher expression of *lmt1* in leaves may be indicative of a role for pinitol in chloroplast osmoprotection, although pinitol is not localized solely to these organelles (Paul and Cockburn, 1989).

The salinity-induced accumulation of *lmt1* is due to transcriptional activation (Figure 3). Based on a number of genes whose patterns of expression have been studied in the ice plant, increased transcription appears to be the primary mechanism of regulation for stress-induced genes (Figure 3; Vernon *et al.*, 1991; Cushman *et al.*, 1989). This is true for both CAM-related mRNAs and those more directly involved in salt tolerance, such as *lmt1*. An understanding of how the plant's transcriptional machinery is cued by the environment is central to an understanding of environmental stress tolerance, and this topic is currently being investigated by the in-depth analysis of selected promoters.

The transcriptional induction of a myo-inositol hydroxymethyl transferase in this facultative halophyte suggests that sugar alcohols play an important role in osmotic stress tolerance. Whether the exact function of these compounds is to serve as non-disruptive cytoplasmic osmolytes or to act as 'osmoprotectants' of protein and membrane structure remains to be elucidated. Experiments utilizing transformation of glycophytes with the *lmt1* gene will address the question of whether production of cyclitols is sufficient for higher plant osmotolerance. If so, the ability of this methyl transferase to create a non-reducing, metabolically inert sugar alcohol from a ubiquitous plant substrate makes it an attractive enzyme for the creation of stress-resistant transgenic plants.

## Materials and methods

### Plant material

*M. crystallinum* were grown from seed as described (Ostrem *et al.*, 1987) and either kept in soil or transferred 14 days after seeding to Hoagland's solution in aerated, light-tight containers for hydroponic growth. Plants were salt stressed when 6 weeks old (unless otherwise noted) with 500 mM NaCl (Ostrem *et al.*, 1987) or, for hydroponics, by addition of 400 mM NaCl to Hoagland's solution. Unstressed controls for each experiment were grown alongside stressed plants and were harvested at the same time. To avoid potential experimental variation due to diurnal fluctuation of RNA and protein levels, all harvests were carried out at the same time during the dark/light cycle (4 h before the end of the light period). Harvested material for RNA extraction was frozen in liquid N<sub>2</sub> and stored at -70°C.

### Subtracted cDNA library construction and screening

Leaf RNA was extracted and poly(A)<sup>+</sup> mRNA prepared as described by Ostrem *et al.* (1987). First-strand cDNAs were generated from poly(A)<sup>+</sup> RNA from stressed plants using a kit (BRL, Gaithersburg, MD). Following first-strand synthesis, RNA was removed by base hydrolysis with 0.1 M NaOH (70°C, 60 min) and single stranded (ss) cDNA was hybridized to a 5.5-fold excess of poly(A)<sup>+</sup> RNA from control plants using the protocol and conditions described (Briehl *et al.*, 1990). Hybridizations were carried out for 18 h. Single-stranded cDNA remaining after hybridization was isolated by differential elution from hydroxyapatite (Davis, 1986) and made double-stranded using cDNA synthesis kit materials (BRL, Gaithersburg, MD) with poly(dN)<sub>6</sub> random primer (Pharmacia Inc., Piscataway, NJ). Double-stranded cDNA was ligated to hemi-phosphorylated *EcoRI*/*NotI* adaptors (Invitrogen Inc., San Diego, CA), ligated into  $\lambda$ ZapII (Stratagene Inc., La Jolla, CA) and packaged using Gigapack *in vitro* packaging extracts (Stratagene Inc., La Jolla, CA). Phage were plated and screened by three

cycles of standard differential plaque hybridization (Maniatis *et al.*, 1982). <sup>32</sup>P-labeled ss cDNA probes used for screening were generated from poly(A)<sup>+</sup> RNA isolated from unstressed *M. crystallinum* and plants salt stressed for 10 h. DNA from selected plaques was rescued as Bluescript SK<sup>-</sup> phagemids (Stratagene Inc., La Jolla, CA) and used to transfect *E. coli* XL1-Blue cells for future DNA analysis and manipulation. Cloned insert DNA was isolated (Maniatis *et al.*, 1982), excised by *Eco*RI digestion and analyzed on agarose gels.

#### RNA isolation, Northern analysis and transcription assays

Total RNA from leaves and roots of hydroponically-grown plants was isolated as previously described (Ostrem *et al.*, 1987), except that the buffer used for root RNA extractions contained additional detergent: 2% SDS and 2% Triton X-100. Total RNA was resolved on formaldehyde-agarose gels and Northern blotted (Thomas, 1983). Blots were probed, washed at high stringency and visualized as previously described (Vernon *et al.*, 1988) using <sup>32</sup>P-labeled, gel-purified *lmt1* insert DNA as a probe. Nuclei isolations from fresh leaf tissue, *in vitro* run-on experiments, DNA slot blotting and hybridization of transcripts to DNA were carried out as described previously (Cushman *et al.*, 1989). The c.p.m. hybridized to filters were directly quantified using a  $\beta$ -scanner (Betagen Inc., Waltham, MA).

#### Gene copy number determination

Nuclear genomic DNA was prepared as described by Bedbrook (1981). 4  $\mu$ g of nuclear DNA was digested to completion with either *Eco*RI, *Hind*III or *Hinc*II, resolved on 1% agarose gels and Southern blotted (Southern, 1975). Blots were processed as described (Vernon *et al.*, 1988). <sup>32</sup>P-labeled probes were generated from gel-purified 5'- or 3'-end restriction fragments of *lmt1* or from full-length insert. Genome copy-number equivalents were based on an estimated *M. crystallinum* haploid genome size of 390 Mbp (De Rocher *et al.*, 1990). Hybridization signals were quantified using a  $\beta$ -scope (Betagen, Inc., Waltham, MA).

#### Sequence analysis

The nucleotide sequence of both strands of the *lmt1* cDNA was obtained by dideoxy sequencing (Sanger *et al.*, 1977) of suitable overlapping *lmt1* restriction fragments and full-length insert cloned into Bluescript KS<sup>+</sup> and KS<sup>-</sup> cloning vectors (Stratagene Inc., La Jolla, CA). IMT1 amino acid sequence and restriction sites were identified using the Mount-Conrad-Meyers program (Williams, 1988). Sequence comparisons to the NBRF protein sequence database were performed using the FASTA program (Pearson and Lipman, 1988).

#### Expression of IMT1 in *E. coli*

The full IMT1 reading frame was cloned into Bluescript KS<sup>+</sup> as a transcriptional fusion in both orientations behind the T7 polymerase promoter. Constructs were transformed into BL21(DE3) cells (Studier *et al.*, 1987), which contain the gene encoding the T7 polymerase under the control of an isopropyl- $\beta$ -D-thiogalactoside (IPTG)-inducible promoter. A fortuitously located AAGAG sequence in the 5' leader of the cDNA was predicted to act as an *E. coli* ribosome binding site. IPTG was added to cultures at a final concentration of 0.5 mM 4 h before harvest. Protein expression was analyzed by SDS-PAGE on 10% acrylamide gels (Ostrem *et al.*, 1987). Samples were prepared by boiling aliquots of *E. coli* cultures for 3 min in an equal volume of SDS extraction buffer.

*In vitro* transcription of the *lmt1* cDNA was carried out with T7 RNA polymerase (Stratagene Inc., La Jolla, CA). *In vitro* translation of the resulting transcript was accomplished using a reticulocyte lysate system (BRL Inc., Gaithersburg, MD). [<sup>35</sup>S]methionine-labeled translation product was resolved by SDS-PAGE alongside *E. coli* protein samples and visualized by fluorography.

#### Soluble protein extraction and activity determination

Soluble protein from *E. coli* transformed with the IMT1 expression construct or native Bluescript KS<sup>+</sup> was extracted from 20–200 ml cultures pelleted at 2500 g for 10 min and resuspended in methyl transferase extraction buffer (MTEB): 100 mM Tris-Cl pH 8, 10 mM EDTA and  $\beta$ -mercaptoethanol; 1 ml/20 ml of culture. Cells were lysed by sonication and extracts were clarified by centrifugation at 10 000 g for 20 min. Total protein concentration was determined by the method of Ghosh *et al.* (1988). Supernatants were either used immediately for assays or stored at -70°C with 5% glycerol.

Myo-inositol O-methyl transferase assays on *E. coli* extracts were carried out in a 200  $\mu$ l volume containing 1.0 mg total soluble protein, 50 mM Tris-Cl pH 8.0, 10 mM MgCl<sub>2</sub> and 1.0 mM myo-inositol. Assays were pre-incubated at 30°C for 5 min and initiated by the addition of S-adenosyl-L-methionine (SAM) to a final concentration of 0.5 mM. SAM stock solution contained unlabeled SAM (Sigma Chemicals St Louis, MO) and <sup>14</sup>C-

labeled SAM (ICN Biochemicals, Irvine, CA) at a 50:1 ratio. Assays were carried out at 30°C for 60 min and terminated by transfer to ice and chloroform extraction. The aqueous phase was subjected to further processing and HPLC analysis (see below).

Total soluble plant protein was extracted as described by Ostrem *et al.* (1987) using MTEB plus 2 mM leupeptin as extraction buffer. Assays were carried out on crude leaf extracts as with *E. coli* extracts, except that myo-inositol was included at a final concentration of 5 mM and assays were run for 2 h. Assays were terminated with phenol-chloroform (1:1) extraction followed by a chloroform extraction.

#### HPLC analysis of assay products

Samples were prepared for HPLC by extraction with 2 vol methanol/chloroform/water (12:5:3) followed by addition of 0.4 ml water. A desalting column of AG50WX4 (Bio-Rad, Richmond, CA) and Amberlite IRA-68 (Sigma Chemicals, St Louis, MO) in OH-form was used to desalt extracts and remove charged species (including remaining [<sup>14</sup>C]SAM). Samples were dried, dissolved in deionized water and filtered through a nylon Acrodisc 13 (Gelman, Ann Arbor, MI). Equal amounts of dissolved carbohydrates from each assay were resolved on a 300  $\times$  7.8 mm HPX87C calcium-form ligand exchange column (Bio-Rad, Richmond, CA) at 85°C with a 0.6 ml/min flow rate using degassed, deionized water as an eluent. Post-column NaOH was added at 0.3 M, 0.6 ml/min, and traces were obtained using a pulsed amperometric detector (Dionex, Sunnyvale, CA) at 35°C and a Spectrophysics SP4290 integrator (Spectrophysics Analytical, San Jose, CA). Fractions were collected at 7.5 s or 0.5 min intervals and scintillation counted.

#### Acknowledgements

We would like to thank Frank Flomerfelt for protocols and advice on subtractive hybridization, Sabine Trinkl for help with sequencing and Dr Elizabeth Vierling, Dr James A. Ostrem and E.J. De Rocher for critical reading of the manuscript. Thanks also to Dr Mitch Tarczynski for helpful comments, discussions and stubborn arguments. We are especially grateful to Pat Adams and Dr Richard Jensen for their expertise in HPLC. Supported by the USDA Environmental Stress Program and the University of Arizona Agricultural Experimental Station.

#### References

- Ahmad, I., Larher, F. and Stewart, G. (1979) *New Phytol.*, **82**, 671–678.
- Bartels, D., Schneider, K., Terstappen, G., Piatowski, P. and Salamini, F. (1990) *Planta*, **181**, 27–34.
- Bartels, D., Engelhardt, K., Rancarati, R., Schneider, K., Rotter, M. and Salamini, F. (1991) *EMBO J.*, **10**, 1037–1043.
- Bedbrook, J. (1981) *Plant Mol. Biol. Newslett.*, **2**, 24.
- Bielecki, R.L. (1982) In Loewus, F.A. and Tanner, W. (eds), *Encyclopedia of Plant Physiology: Plant Carbohydrates I—Intracellular Carbohydrates*. Springer-Verlag, Berlin, Vol. 13A, pp. 158–192.
- Binzel, M.L., Hess, F.D., Bressan, R.A. and Hasegawa, P.M. (1988) *Plant Physiol.*, **86**, 607–614.
- Bohnert, H.J. *et al.* (1988) *Plant Mol. Biol. Rep.*, **6**, 10–28.
- Boyer, J.S. (1982) *Science*, **218**, 443–448.
- Briehl, M.M., Flomerfelt, F.A., Wu, X. and Miesfeld, R.L. (1990) *Mol. Endocrinol.*, **4**, 287–294.
- Bugos, R.C., Chiang, V.L.C. and Campbell, W.H. (1991) *Plant Mol. Biol.*, **17**, 1203–1215.
- Claes, B., Dekeyser, R., Villaroel, R., Van den Buicke, M., Bauw, G., Van Montague, M. and Caplan, C. (1990) *Plant Cell*, **2**, 19–27.
- Cohen, A., Plant, A.L., Moses, M.S. and Bray, E.A. (1991) *Plant Physiol.*, **97**, 1367–1374.
- Coutinho, C., Bernardes, E., Felix, D. and Panek, A.D. (1988) *J. Biotechnol.*, **7**, 23–32.
- Csonka, L.N. and Hanson, A.D. (1991) *Annu. Rev. Microbiol.*, **45**, 569–606.
- Cushman, J.C., Meyer, G., Michalowski, C., Schmitt, J.M. and Bohnert, H.J. (1989) *Plant Cell*, **1**, 715–725.
- Davis, M.M. (1986) *Handbook Exp. Immunol.*, **76**, 1–13.
- De Rocher, E.J., Harkins, K.R., Galbraith, D.W. and Bohnert, H.J. (1990) *Science*, **250**, 99–101.
- Dittrich, P. and Brandl, A. (1987) *Phytochemistry*, **26**, 1925–1926.
- Dittrich, P. and Korak, A. (1984) *Phytochemistry*, **23**, 65–66.
- Dure, L., III, Crouch, M., Harada, J., Ho, T.-H.D., Mundy, J., Quatrano, R., Thomas, T. and Sung, Z.R. (1989) *Plant Mol. Biol.*, **12**, 475–486.

- Ghosh, S., Gepstein, S., Heikkilä, J.J. and Dumbroff, E.B. (1988) *Anal. Biochem.*, **169**, 227–233.
- Gomez, J., Sanchez-Martinez, D., Steifel, V., Rigau, J., Puigdomenech, P. and Pages, M. (1988) *Nature*, **334**, 262–264.
- Gorham, J., Hughes, L.I. and Wyn Jones, R.G. (1981) *Physiol. Plant.*, **53**, 27–33.
- Gorham, J., Tomar, O. and Wyn Jones, R.G. (1988) *J. Plant Physiol.*, **132**, 678–682.
- Gowri, G., Bugos, R., Campbell, W.H., Maxwell, C. and Dixon, R. (1991) *Plant Physiol.*, **97**, 7–14.
- Hellebus, J.A. (1976) *Annu. Rev. Plant Physiol.*, **27**, 485–505.
- Hoffmann-Ostenhoff, O., Pittner, F. and Koller, F. (1978) In Wells, W.W. and Eisenberg, F. (eds), *Cyclitols and Phosphoinositides*. Academic Press, New York, pp. 233–247.
- Iraki, N.M., Bressan, R.A., Hasegawa, P.M. and Carpita, N.C. (1989) *Plant Physiol.*, **91**, 39–47.
- Ishida, I., Obinata, M. and Deguchi, T. (1987) *J. Biol. Chem.*, **262**, 2895–2899.
- Jefferies, R.L. (1981) *Bioscience*, **31**, 42–46.
- Laurie, S. and Stewart, G. (1990) *J. Exp. Bot.*, **41**, 1415–1422.
- Loewus, F. and Dickinson, D.B. (1982) In Loewus, F.A. and Tanner, W. (eds) *Encyclopedia of Plant Physiology: Plant Carbohydrates I: Intracellular Carbohydrates*. Springer-Verlag, Berlin, Vol. 13A, pp. 193–206.
- Loewus, F. and Loewus, M. (1980) In Stumpf, P. and Conn, E. (eds), *The Biochemistry of Plants: Carbohydrate Structure*. Academic Press, New York, Vol. 3, pp. 43–76.
- Maniatis, T., Fritsch, E. and Sambrook, J. (1982) *Molecular Cloning, A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Mason, H. and Mullet, J. (1990) *Plant Cell*, **2**, 569–579.
- McCue, K.F. and Hanson, A.D. (1990) *Trends Biotechnol.*, **8**, 358–362.
- Mundy, J. and Chua, N.-H. (1988) *EMBO J.*, **7**, 2279–2286.
- Nguyen, A. and Lamant, A. (1988) *Phytochemistry*, **27**, 3423–3427.
- Ostrem, J.A., Vernon, D.M. and Bohnert, H.J. (1990) *J. Biol. Chem.*, **265**, 3497–3502.
- Ostrem, J.A., Olsen, S.W., Schmitt, J.M. and Bohnert, H.J. (1987) *Plant Physiol.*, **84**, 1270–1275.
- Paul, M.J. and Cockburn, W. (1989) *J. Exp. Bot.*, **40**, 1093–1098.
- Pearson, W.R. and Lipman, D.J. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2444–2448.
- Plant, A.L., Cohen, A., Moses, M.S. and Bray, E.A. (1991) *Plant Physiol.*, **97**, 900–906.
- Pollard, A. and Wyn Jones, R.G. (1979) *Planta*, **144**, 291–298.
- Popp, M. (1984) *Z. Pflanzenphysiol.*, **113**, 411–421.
- Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
- Schmitt, J.M., Michalowski, C.M. and Bohnert, H.J. (1988) *Photosynth. Res.*, **17**, 159–171.
- Schobert, B. (1977) *J. Theor. Biol.*, **68**, 17–26.
- Singh, N., Nelson, D., Kuhn, D., Hasegawa, P.M. and Bressan, R.A. (1989) *Plant Physiol.*, **90**, 1096–1101.
- Skriver, K. and Mundy, J. (1990) *Plant Cell*, **2**, 503–512.
- Southern, E.M. (1975) *J. Mol. Biol.*, **98**, 503–517.
- Studier, F.W., Rosenberg, A.H., Dunn, J.J. and Dubendorff, J.W. (1987) *Methods Enzymol.*, **185**, 60–89.
- Thomas, P.S. (1983) *Methods Enzymol.*, **100**, 255–266.
- Ting, I.P. (1985) *Annu. Rev. Plant Physiol.*, **36**, 595–622.
- Vernon, D.M., Ostrem, J.A., Schmitt, J.M. and Bohnert, H.J. (1988) *Plant Physiol.*, **86**, 1002–1004.
- Vernon, D.M., Ostrem, J.A. and Bohnert, H.J. (1991) *Plant Physiol.*, **96**, abs. no. 99.
- Williams, K.M. (1988) *Comput. Appl. Biosci.*, **4**, 211.
- Winter, K. and Gademann, R. (1991) *Plant Physiol.*, **95**, 768–776.
- Winter, K. and von Willert, D.J. (1972) *Z. Pflanzenphysiol.*, **67**, 166–170.
- Wyn Jones, R.G. and Gorham, J. (1983) In Lange, O.L., Nobel, P.S., Osmond, C.B. and Ziegler, H. (eds), *Encyclopedia of Plant Physiology: Responses to the Environment*. Springer-Verlag, Berlin, Vol. 12C, pp. 35–58.
- Yancey, P., Lark, M., Hand, S., Bowler, R. and Somero, G. (1982) *Science*, **217**, 1214–1222.

Received on February 25, 1992

## Common amino acid sequence domains among the LEA proteins of higher plants

Leon Dure III,<sup>1\*</sup> Martha Crouch,<sup>2</sup> John Harada,<sup>3</sup> Tuan-Hua David Ho,<sup>4</sup> John Mundy,<sup>5</sup> Ralph Quatrano,<sup>6</sup> Terry Thomas<sup>7</sup> and Z. R. Sung<sup>8</sup>

<sup>1</sup>Department of Biochemistry, University of Georgia, Athens, GA 30602, USA (\*author for correspondence); <sup>2</sup>Department of Biology, Indiana University, Bloomington, IN 47405, USA; <sup>3</sup>Department of Botany, University of California, Davis, CA 95616, USA; <sup>4</sup>Biology Department, Washington University, St. Louis, MO 63130, USA; <sup>5</sup>Laboratory of Plant Molecular Biology, The Rockefeller University, 1230 York Avenue, New York, NY 10021, USA; <sup>6</sup>Department of Biology, University of North Carolina, Chapel Hill, NC 27549, USA; <sup>7</sup>Department of Biology, Texas A & M University, College Station, TX 77843, USA; <sup>8</sup>Departments of Genetics and of Plant Pathology, University of California, Berkeley, CA 94720, USA

Received 12 November 1988; accepted in revised form 30 December 1988

**Key words:** LEA protein sequence, plant desiccation, abscisic acid

### Abstract

LEA proteins are late embryogenesis abundant in the seeds of many higher plants and are probably universal in occurrence in plant seeds. LEA mRNAs and proteins can be induced to appear at other stages in the plant's life by desiccation stress and/or treatment with the plant hormone abscisic acid (ABA). A role in protecting plant structures during water loss is likely for these proteins, with ABA functioning in the stress transduction process. Presented here are conserved tracts of amino acid sequence among LEA proteins from several species that may represent domains functionally important in desiccation protection. Curiously, an 11 amino acid sequence motif is found tandemly repeated in a group of LEA proteins of vastly different sizes. Analysis of this motif suggests that it exists as an amphiphilic  $\alpha$  helix which may serve as the basis for higher order structure.

### Introduction

The amino acid sequences for a number of LEA proteins have been deduced in the past year from cDNA and/or genomic DNA sequences. These include 6 proteins from cotton [1], 2 for carrot [3, Seffens and Thomas, in preparation] and 1 each from barley [8], rice [12], rape [7] and wheat [10]. All of these proteins and their mRNAs accumulate in embryo tissues of seeds as they

approach maturity and begin to desiccate. The LEA mRNAs are not readily detected in leaves or roots of non-stressed plants nor in young immature seed tissues. The cotton LEA mRNAs can be induced to high levels prematurely by treating immature embryos with ABA [6]. The mRNA for the LEA protein of barley, rice and wheat can be induced to reappear in germinating seedlings by treatment with ABA [8, 12, 2]. Severe desiccation of leaves induces the rape LEA mRNA [7]. Thus

Table 1. LEA proteins of known sequence.

Species	Derivation	Amino acids	Mol. wt.	Clone name	Ref.
<i>Group 1</i>					
wheat	complete cDNA	93	9945	Em protein	[10]
cotton	complete cDNA	120	13 150	D 19	[1]
<i>Group 2</i>					
rice	genomic DNA	163	16 526	rab 21	[12]
cotton	complete cDNA	145	15 793	D 11	[1]
<i>Group 3</i>					
cotton	composite cDNA/ genomic DNA	136	14 597	D 7	[1]
carrot (I)	complete cDNA	163	16 989	Dc-3	[+]
barley	complete cDNA	213	21 803	pHVal	[8]
rape	complete cDNA	280	30 360	pLEA 76	[7]
carrot (II)	incomplete cDNA	323*	66 000**	Dc-8	[3]

[+] Seffens and Thomas, in preparation

\* Of sequenced fragment.

\*\* From gel electrophoresis.

there may be a functional linkage between desiccation stress, ABA and LEA proteins. The amino acid sequences of these LEA proteins from the different species have been compared by dot matrix analysis and three of the cotton LEA proteins were found to share regions of homology with proteins from the other plant families. These groupings are: cotton LEA D 19 and the wheat Em protein; cotton LEA D 11 and the rice protein rab 21; cotton LEA D 7 and a LEA protein from barley, rape and two from carrot. These groups are listed in Table 1. The carrot sequences, although sharing regions of amino acid homology, represent 2 distinctly different proteins and are referred to as carrot I and II. The regions of homology among the Group 3 LEA proteins involve a repeating tract of 11 amino acids that possibly exists in an  $\alpha$  helix with twisting amphiphilic surfaces.

### Dot matrix analysis

The sequence comparisons were carried out using the computer program DSAS 2 created by J. Neigel, Southwestern Louisiana University. The stringency for the comparisons between different

proteins required that 7 of 11 amino acids match to generate a dot, but allowed synonym amino acids (having functionally similar R groups) to be considered identical. These groupings are: Asp = Glu, Asn = Gln, Arg = Lys, Ser = Thr, Phe = Tyr = Trp, Ala = Val = Ile = Leu = Met. These groupings reduce the number of amino acids from 20 to 10, but requiring a 7 of 11 match predicts a random match only once in over 200 000 11 amino acid comparisons. Thus random background dots are extremely rare.

Such a calculation of background random dots assumes that amino acid of usage is random, which is not true. However, the tracts of sequence homology observed in these comparisons are clearly meaningful, and denote the conservation of ancestral sequence motifs.

Where a protein sequence is compared with itself to visualize repeated amino acid sequences, the stringency was raised to require an 8 of 11 match.

Tracts of nucleotide sequence homology between cDNAs/genomic DNAs for these proteins is not readily discernible in all cases by dot matrix analysis. In order to generate diagonal tracts of dots, the stringency of the comparisons must be lowered to the point that considerable

random background dots are generated. In the two instances where enough homology exists to allow background to be reduced, the generation of a dot required that 9 of 12 nucleotide match. This predicts the likelihood of a random background dot to be roughly 1 in 3000 twelve nucleotide comparisons.

### Cotton/wheat LEA proteins

Dot matrix amino acid sequence comparison between the cotton LEA D 19 protein and the Em protein of wheat is shown in Fig 1 (top right). It is apparent from this figure that these proteins have not diverged greatly in the > 100 million years that separates the evolution of cotton and wheat. In Fig. 2 (top) the sequences of these two proteins are aligned. The cotton protein is seen to have a few additional amino acids near the N terminus and an additional 18 amino acids at the

C terminus not found in the wheat protein. Of the 93 amino acids aligned, 65 are identical. Of the 28 that are different, 9 involve differences between synonyms. Neither protein contains Cys or Trp. No structural domains such as regions of amphiphilic  $\alpha$  helices are suggested by the sequences of these two proteins. Fig. 1 (top left) compares the nucleotide sequences of the cDNAs from which the amino acid sequences were derived. Fragments of a diagonal line of homology are visible, indicating that high nucleotide homology is limited to a few scattered regions.

### Cotton/rice LEA proteins

Fig. 1 (bottom right and left) gives the dot matrix comparison of amino acids and protein coding nucleotides respectively for the cotton LEA protein D 11 and the rab 21 protein from rice embryos. It is apparent from these plots that

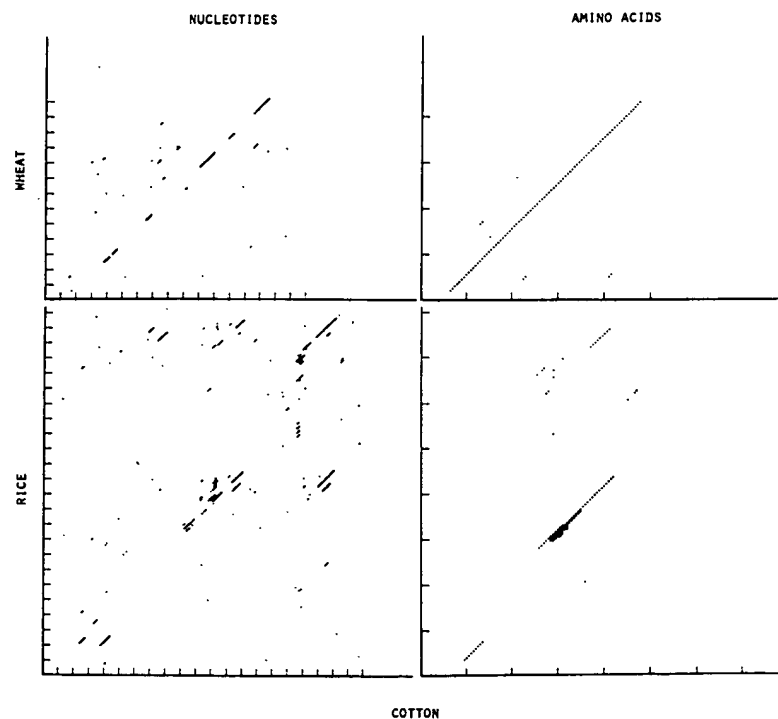


Fig. 1. Dot matrix analyses of amino acid sequences (right) and nucleotide sequences (left) of the Group 1 (top) and 2 (bottom) LEA proteins. Tic marks along the x and y axes represent 20 amino acids or nucleotides. In the amino acid comparison, a match of 7 in a span of 11 yields a dot. Functionally similar amino acids are treated as identical (see text). In the nucleotide comparisons, a match of 9 in a span of 12 yields a dot.

## COMPARISON OF THE WHEAT LEA PROTEIN WITH THE COTTON D 19 PROTEIN

cotton	1	MET	ALA	SER	GLU	GLN	TYR	GLN	ALA	MET	ARG	ASN	ALA	PRO	GLN	GLU	GLU	LYS	GLU	GLU	LEU
wheat	1				gly	...	...	...	...	...	...	...	...	...	...	...	...	ARG	ser	gln	
	21	ASP	ALA	ARG	ALA	LYS	GLN	GLY	GLU	THR	VAL	VAL	PRO	GLY	GLY	THR	ARG	GLY	LYS	SER	LEU
	12		arg	LYS		ARG	glu										gly				
	41	ASP	ALA	GLN	ILE	ASN	LEU	ALA	GLU	GLY	ARG	HIS	LYS	GLY	GLY	GLU	THR	ARG	LYS	GLN	GLN
	32	GLU			glu						ser	ARG				gln		ARG	glu		
	61	LEU	GLY	THR	GLU	GLU	TYR	GLN	GLU	MET	GLY	ARG	LYS	GLY	GLY	LEU	SER	ASN	SER	ASP	MET
	52	MET		glu			ser	gln									thr	asn		glu	
	81	SER	GLY	GLY	GLU	ARG	ALA	ALA	ASP	GLU	GLY	VAL	THR	ILE	ASP	GLU	SER	LYS	PHE	ARG	THR
	72				ASP			arg				ILE	asp						LYS		
	101	LYS	LYS	LEU	ASN	ILE	ILE	SER	HIS	ASP	ARG	THR	LEU	ALA	THR	PRO	PHE	ILE	TYR	LYS	PRO
	92		ser	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

## COMPARISON OF THE RICE LEA PROTEIN WITH THE COTTON D 11 PROTEIN

rice	1	MET	GLU	HIS	...	GLN	GLY	GLN	HIS	GLY	HIS	VAL	...	...	...	...	...	...	...	...	...
cotton	1	MET	ala	HIS	phe	GLN	asn	GLN	tyr	ser	ala	pro	glu	val	thr	gln	thr	asp	ala	tyr	gln
	11	...	...	[THR	SER	ARG	VAL	ASP	GLU	TYR	GLY	ASN	PRO	VAL	GLY	THR]	GLY	ALA	GLY	HIS	ALA
	21	asn	pro	[THR	arg	ARG	thr	ASP	GLU	TYR	GLY	ASN	PRO	ILE	pro	THR]	gln	glu	thr	gly	arg
				*	*		*	*	*	*	*	*	*	*	*	*					
	29	GLN	MET	GLY	THR	ALA	GLY	MET	GLY	THR	HIS	GLY	THR	ALA	GLY	THR	GLY	ARG	GLN	PHE	GLN
	41	gly	ILE	ieu	gly	ILE	GLY	gly	his	his	HIS	GLY	gly	his	his	gly	leu	his	arg	thr	gly
	49	VAL	LEU	GLN	ARG	[SER	GLY	SER	SER	SER	SER	SER	SER	SER	SER	GLU	ASP	ASP	GLY	MET	GLY
	61	...	...	...	...	[SER	SER	SER	SER	SER	SER	SER	SER	SER	SER	GLU	ASP	GLU	GLY	thr	GLY
						*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
	69	ARG	ARG	LYS	LYS	GLY	ILE	LYS	GLU	LYS	ILE	LYS	GLU	LYS	LEU	PRO	GLY	GLY	ASN	LYS]	GLY
	76	LYS	LYS	LYS	LYS	GLY	LEU	LYS	GLU	ARG	LEU	LYS	GLU	LYS	ILE	PRO	GLY	...	ASN	LYS]	glu
	89	GLU	GLN	GLN	HIS	ALA	MET	GLY	GLY	THR	GLY	THR	GLY	THR	GLY	THR	GLY	THR	GLY	THR	GLY
		...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	109	GLY	ALA	TYR	GLY	GLN	GLN	GLY	[HIS	GLY	THR	GLY	MET	THR	THR	GLY	THR	THR]	GLY	ALA	HIS
	95	...	...	...	...	...	...	...	[HIS	gln	SER	gln	ALA	THR	SER	thr	THR	THR]	pro	gly	gln
									*	*	*	*	*	*	*	*	*	*	*	*	*
	129	GLY	THR	THR	THR	ASP	THR	GLY	GLU	LYS	LYS	GLY	ILE	MET	ASP	LYS	ILE	LYS	GLU	LYS	
	108	GLY	pro	THR	tyr	his	gln	his	his	arg	glu	glu	arg	ser	asp	gly	gln	gly	glu	ala	pro
	149	LEU	PRO	GLY	GLN	HIS	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	128	trp	ser	pro	GLN	pro	leu	ile	ser	cys	leu	trp	ser	ala	ile	ser	tyr				

Fig. 2. Sequence alignment of the Group 1 (top) and 2 (bottom) LEA proteins. Missing amino acids (gaps) are indicated with dots. In the top alignment, amino acids in the wheat protein that differ from those of the cotton protein appear in lower case letters. Those that differ but represent synonym functions appear in upper case letters. In the bottom alignment identical or synonym amino acids appear in upper case letters, mismatched amino acids in lower case letters. Region of high homology are bracketed, sequences repeated in the cotton protein are underscored with asterisks, those repeated in the rice protein are overscored with asterisks. The tract of contiguous Ser residues is overlined.

homologous tracts exist only in portions of these proteins. In Fig. 2 (bottom) these amino acid sequences have been aligned by eye. The regions of homology, seen as diagonals in Fig. 1, are bracketed. The cotton protein has a short se-

quence occurring twice (6 of 7 amino acids) near the N terminus that is underscored by asterisks in the figure. It appears once in the rice protein as part of the first region of homology. The second region of homology is curious in that it begins with

a tract of 9 contiguous Ser residues in the cotton protein, 8 of 9 in the rice protein (overlined in the figure). Also within this second region of homology is a stretch of 14 amino acids that occurs again in the rice protein, overscored with asterisks. Both proteins are very high in Gly and hydroxyl-bearing amino acids, the rice protein containing a tract of 6 contiguous Thr-Gly pairs. The high Gly content would seem to preclude extensive  $\alpha$  helical domains in either protein.

### Cotton/barley/carrot/rape LEA proteins

Five other LEA proteins from 4 plant species (cotton, carrot, barley and rape) ranging in size from roughly 14.5 to over 60 kD have regions of sequence homology. Each of these proteins contain repeated amino acid tracts. This is shown in Fig. 3 (left column) which gives the dot matrix analysis of each protein compared with itself. The repeated sequences occur principally in the N terminal 1/2 of these proteins. The carrot II sequence used here is from an incomplete cDNA sequence; further data suggest that it represents the N terminal half of the carrot protein as well (Franz *et al.*, in preparation). The number of the repeats in each protein varies considerably as does the size of the proteins. The pair wise sequence comparison of all 5 of these proteins, shown in the remainder of Fig. 3, indicates that the repeated sequences are shared among these proteins. Regions of homology observed between the carrot II protein and the cotton, rape and even the carrot I protein are very limited, but, as will be seen, the repeated sequence of the carrot II protein has diverged considerably from the consensus repeated sequence found in the other proteins.

Figure 4 gives the amino acid sequence of the regions carrying the repeat in the 5 proteins beginning with the smallest region, containing 5 repeat units, which is found in the cotton D 7 protein. The repeating unit is seen to be a tract of 11 amino acids that is repeated contiguously. Clearly there is considerable degeneracy in the repeat within a single protein as well as between proteins. In two instances (barley and carrot I

protein) the continuity of the repeats is interrupted by tracts seemingly unrelated to the repeat sequences. Curiously, in 2 of the 3 such instances, the intervening tract is 11 amino acids. Further, amino acids of the repeating unit in some instances have been lost, creating gaps in the unit (barley, rape and carrot II).

Scrutiny of the repeating units of the rape and carrot II proteins suggests that during the evolution of these proteins the original 11-mer repeat unit was duplicated once, followed by divergence of each of the 2 units. Subsequently the 22 amino acid unit was duplicated several times, i.e., two somewhat different 11-mer repeats can be discerned in these proteins that occur in an alternating pattern. Thus these sequences are presented in Fig. 4 in tracts of 22, spaced to indicate the 2 11-mer units.

Despite the degeneracies listed above, it seems likely that these 5 proteins have a common heritage based upon the 11-mer motif and that this conserved motif is important in function.

The salient chemical features of the 11-mer motif become apparent when the consensus chemical property of the amino acids occurring at each of the 11 positions is determined by summation which is presented in Fig. 5. Since the 11-mer motif in the carrot II protein is divergent from the other 4, it is treated independently in this summation. The summation shows that in positions 1, 2, 4, 5 and 9, Ala and Thr predominate. Since these 2 amino acids appear interchangeable in these positions, the methyl function of these residues must be required. Positions 3, 7 and 11 are occupied principally by either a negatively charged residue or Gln and thus can be considered to require a very polar R group. Positions 6 and 8 prefer positively charged residues. In position 10 there seems to be no function required, 12 different amino acids appearing in the 31 repeat units examined. From this, the functional consensus shown in Fig. 5 was arrived at for these 4 LEA proteins and the frequency of occurrence in each position given in the figure.

The carrot II protein summation reveals that the functional consensus for positions 2, 5, 6, 7, 9 and 11 are the same as for other LEA proteins



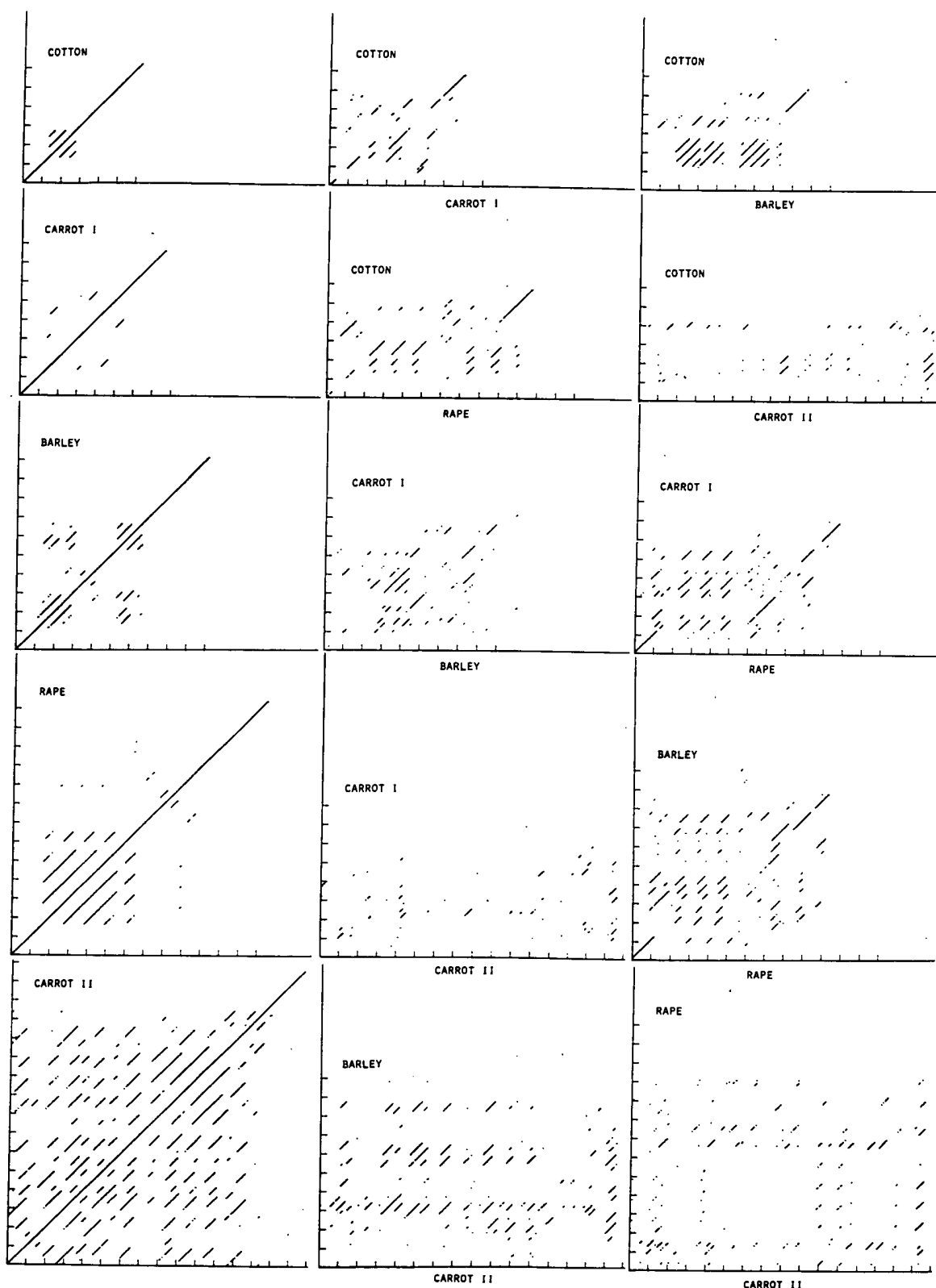


Fig. 3. Dot matrix analyses of Group 3 LEA proteins. In the left column the amino acid sequence of each protein is compared with itself, a match of 8 of 11 residues yields a dot. In the pair wise comparisons 7 of 11 residues must match to yield a dot and synonym amino acids are considered identical (see text). Tic marks represent 20 amino acids.

# REPEATED MOTIF IN LEA PROTEINS FROM COTTON, CARROT, BARLEY AND RAPE

## cotton D 7 Lea

31 LYS ALA GLU ALA ALA LYS GLN LYS THR MET GLU  
 42 THR ALA GLU ALA ALA LYS GLN LYS THR MET GLU  
 53 THR ALA GLU ALA ALA LYS GLN LYS THR ARG GLY  
 64 ALA ALA GLU THR THR ASN ASP LYS THR LYS GLN  
 75 THR ALA GLY ALA ALA ARG GLY LYS ALA GLU GLU

## carrot I Lea

32 LYS ALA GLN ALA ALA LYS ASP LYS ALA SER GLU  
 43 MET ALA GLY SER ALA ARG ASP ARG THR VAL GLU ser lys asp gln thr gly ser tyr val ser asp  
 65 LYS ALA GLY ALA VAL LYS ASP LYS THR CYS GLU  
 76 THR ALA GLN ALA ALA LYS GLU LYS THR GLY GLY  
 87 ALA MET GLN ALA THR LYS GLU LYS ALA SER GLU met gly glu ser ala lys glu  
 105 THR ALA VAL ALA GLY LYS GLU LYS THR GLY GLY

## barley Lea

26 MET MET GLY ALA THR LYS GLN GLY ALA GLY GLN  
 37 THR THR GLU ALA THR LYS GLN LYS ALA GLY GLU  
 48 THR ALA GLU ALA THR LYS GLN LYS THR GLY GLU  
 59 THR ALA GLU ALA ALA LYS GLN LYS ALA ALA GLU  
 70 ALA LYS ASP LYS THR ALA GLN  
 77 THR ALA GLN ALA ALA LYS ASP LYS THR TYR GLU  
 88 THR ALA GLN ALA ALA LYS GLU ARG ALA ALA GLN gly lys asp gln thr gly ser ala leu gly glu  
 101 LYS THR GLU ALA ALA LYS GLN LYS ALA ALA GLU  
 121 THR THR GLU ALA ALA LYS GLN LYS ALA ALA GLU  
 132 ALA THR GLU ALA ALA LYS GLN LYS ALA SER ASP

## rape Lea

31 LYS ALA GLU GLU GLY LYS ASP LYS THR SER GLN THR ALA GLN LYS ALA GLN GLN LYS ALA GLN GLU  
 53 THR ALA GLN ALA ALA LYS ASP LYS THR SER GLN ALA ALA GLN THR THR GLN GLN LYS ALA GLN GLU  
 75 THR ALA GLN ALA ALA LYS ASP LYS THR SER GLN ALA ALA GLN THR THR GLN GLN LYS ALA GLN GLU  
 97 THR ALA GLN ALA ALA LYS ASP LYS THR SER GLN ALA ALA GLN THR THR GLN GLN LYS ALA HIS GLU  
 119 THR THR GLN SER SER LYS GLU LYS THR SER GLN ALA ALA GLN THR ALA GLN GLU LYS ALA ARG GLU  
 141 THR LYS ASP LYS THR

## carrot II Lea

2 LYS ALA LYS MET ALA LYS ASP THR THR MET GLY LYS ALA GLY GLU TYR LYS ASP TYR THR ALA GLN  
 24 LYS ALA GLU GLU ALA LYS GLU LYS ALA ALA GLN  
 35 LYS ALA GLU GLU THR LYS GLU LYS ALA GLY GLU TYR LYS ASP TYR THR ALA GLN  
 53 LYS ALA GLY GLU ALA LYS ASP THR THR LEU GLY LYS ALA GLY GLU TYR LYS ASP THR ALA ALA GLN  
 75 LYS ALA ALA GLU ALA LYS ASP THR THR ALA GLN  
 86 LYS ALA ALA GLU ALA LYS GLU LYS THR GLY GLU TYR LYS ASP THR ALA ALA GLN  
 104 LYS ALA ALA GLU ALA LYS VAL LEU ALA ALA GLN  
 115 LYS ALA ALA GLU ALA LYS ASP THR THR GLY LYS ASP GLY GLU TYR LYS ASP TYR ALA ALA GLN  
 136 LYS ALA ALA GLU ALA LYS ASP ALA THR MET GLN LYS PRO VAL SER ASN LYS ASP TYR ALA ALA GLN  
 158 LYS THR ALA GLU THR LYS ASP ALA THR MET GLU LYS ALA LYS GLU TYR LYS GLU TYR ALA ALA GLN  
 180 LYS ALA ALA GLU ALA LYS ASP ALA THR MET GLN LYS THR GLY GLU TYR LYS ASP TYR SER ALA GLN  
 202 LYS ALA ALA GLU THR LYS ASP ALA THR MET GLN LYS THR LYS GLU TYR LYS ASP TYR THR ALA GLN  
 224 LYS ALA ALA GLU THR LYS ASP ALA THR MET GLU  
 235 LYS ALA LYS GLU ALA LYS ASP THR THR VAL GLN LYS THR GLY GLU TYR LYS ASP TYR ALA ALA GLU  
 257 LYS ALA LYS GLU GLY LYS ASP VAL THR VAL GLU  
 268 LYS ALA LYS GLU GLY LYS ASP THR THR VAL GLY

Fig. 4. Amino acid sequences of those regions of the Group 3 LEA proteins that contain the 11-mer repeating unit. Tracts showing no homology with the repeating unit extend to the right and are in lower case letters. Gaps in the repeating unit are left blank.

SUMMATION OF AMINO ACIDS IN REPEATING UNIT											
Position	1	2	3	4	5	6	7	8	9	10	11
(all except carrot II)											
THR 17	ALA 23	GLN 14	ALA 21	ALA 19	LYS 24	GLN 14	LYS 29	THR 17	SER 8	GLU 18	
ALA 7	THR 5	GLU 11	THR 5	THR 8	GLN 5	ASP 11	ARG 2	ALA 15	GLY 5	GLN 9	
LYS 5	MET 2	GLY 4	SER 2	GLY 2	ARG 2	GLU 6	GLY 1		ALA 5	GLY 3	
MET 2		VAL 1	LYS 1	SER 1	ASN 1	GLY 1			GLN 3	ASP 1	
			GLU 1	VAL 1					MET 2		
									ARG 2		
									HIS, LYS		
									GLU, VAL		
									TYR, CYS		
									1 each		
methyl	methyl	polar	methyl	methyl	positive	polar	positive	methyl	X	polar	
24/31	28/30	25/30	26/30	27/31	26/32	31/32	31/32	32/32		28/31	
(carrot II)											
LYS 26	ALA 19	ALA 9	GLU 24	ALA 10	LYS 26	ASP 21	THR 8	THR 15	ALA 13	GLN 16	
	THR 5	GLY 8	SER 1	TYR 9		GLU 4	TYR 8	ALA 8	MET 6	GLU 4	
	ASP 1	LYS 6	MET 1	THR 4		VAL 1	ALA 5	SER 1	VAL 3	GLY 3	
	PRO 1	GLU 2		GLY 2			VAL 1		LEU 1		
		VAL 1		ASN 1			LEU 1		GLY 1		
							LYS 1				
positive	methyl	X	negative	methyl	positive	negative	apolar	methyl	apolar	polar	
26/26	24/26		24/26	14/26	26/26	25/26	23/24	23/24	23/24	20/23	

Fig. 5. The occurrence of amino acids in each position of the 11-mer repeat unit has been collated for 4 of the Group 3 LEA proteins (top half) and for the carrot II LEA protein (bottom half). The consensus chemical function for each position is shown as is the number of times this function appears in this position compared to the total number of repeating units analyzed at this position.

of this group. However, positions 1, 4, and 8 are reversed in hydrophilicity/hydrophobicity from the others. Position 3 in the carrot II protein does not seem to require a specific functional property whereas position 10 clearly requires an apolar moiety.

### Hydrophilicity of the LEA proteins

Figure 6 presents hydropathy plots of all 9 of the LEA proteins calculated according to the amino acid values of Kyte and Doolittle [9]. In the plots of the proteins carrying the 11-mer repeats (right column), the regions containing the repeat are indicated by a horizontal bar. The predominant feature of all these plots is the overriding hydrophilicity of the proteins. The single region in the carrot II partial protein that is hydrophobic is due to the sequence Val-Leu-Ala-Ala occurring in one of the repeating units (underlined in Fig. 4). Although the calculation of the hydrophilicity of

a span of amino acids based on the additivity of the properties of free amino acids may be questioned [13], the virtually unbroken hydrophilicity of these proteins makes it unlikely that they span membranes or form globular structures with buried hydrophobic cores. The fact that all the cotton LEA proteins (identified by hybrid-arrested translation and 2D electrophoresis [6]) are easily solubilized by homogenizing tissue in 0.1 M NaCl pH 7.0 [5] predicts that they are not elements of structure nor compartmentalized in cells.

### Possible structure of the repeating unit

The multiple and contiguous occurrence of a tract of amino acids with a periodicity of 11 suggests that this arrangement allows for the formation of a secondary and tertiary structure that may be necessary for protein function. An obvious possibility is the  $\alpha$  helix, perhaps with amphiphilic

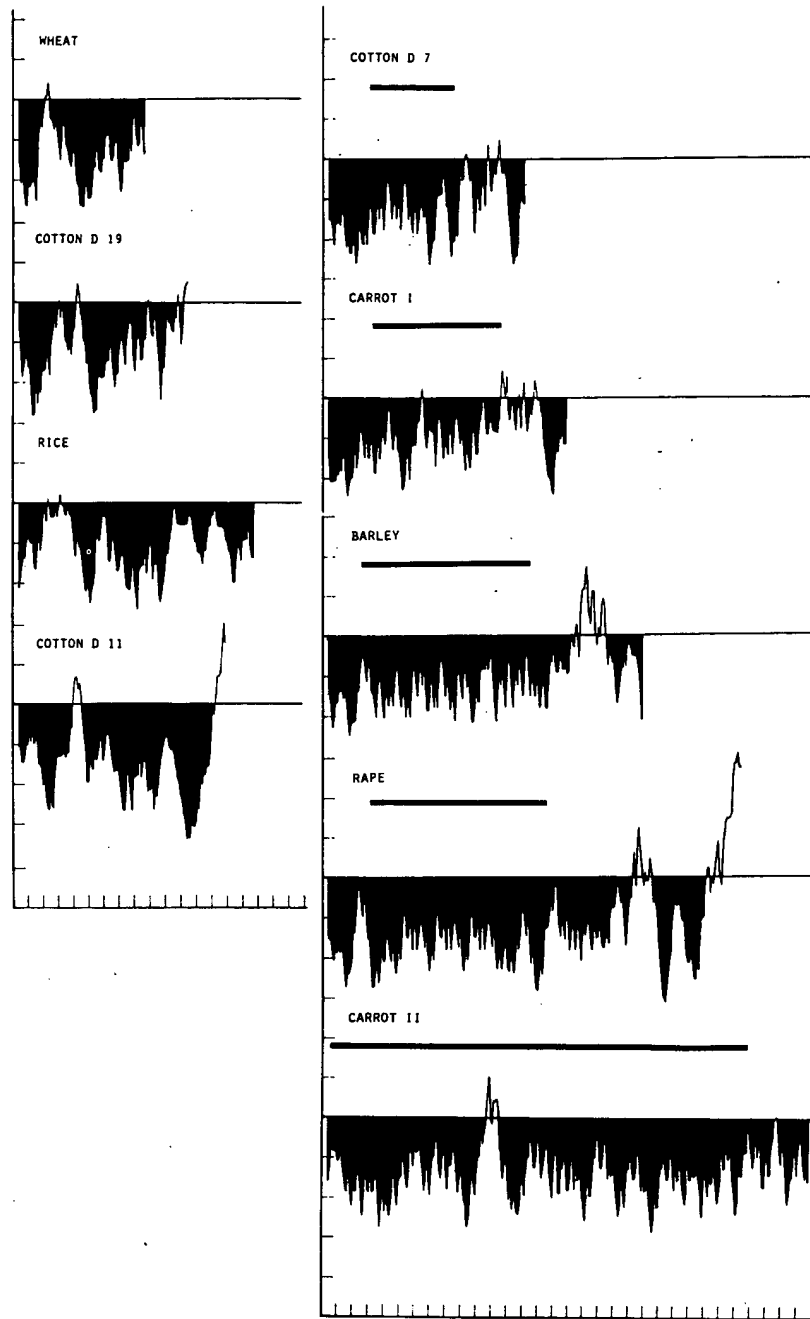


Fig. 6. Hydropathy analyses of the LEA proteins calculated as given in Baker *et al.* [1]. Tic marks along the left border represent 9 hydropathy units. Those values below the zero line are negative, shaded and hydrophilic. The bottom tic marks represent 10 amino acids. The amino acid span for each value is 9. Groups 1 and 2 LEA proteins are in the left column; Group 3 in the right column where regions of the proteins carrying the 11-mer repeating unit are indicated with a bar.

properties. The absence of Pro, the paucity of Gly and the interspersion of polar and apolar residues in the repeat unit point to this possibility. Figure 7 shows the 'helical wheel' diagram [14] for an  $\alpha$  helix of 54 amino acids (15 turns) composed of 5 repeating units of the 11-mer motif consensus for 4 of this group of LEA proteins (A) and consensus for the carrot II LEA protein (B). In both wheels the apolar residues are shaded, the polar or charged residues indicated and the positions of no apparent specificity left blank. In both cases

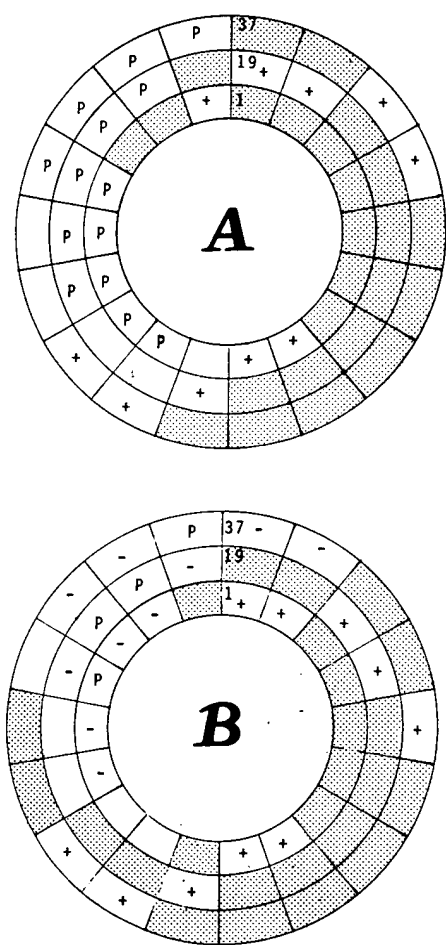


Fig. 7. Helical wheel diagrams of the 11-mer repeating unit found in the Group 3 LEA proteins. 54 amino acids (15 turns of the  $\alpha$  helix) representing 5 contiguous repeating units are shown. The chemical functions of the amino acid R groups are shown as: apolar, shaded; positive and negative as indicated; unspecified positions left blank. A represents the consensus repeating unit for 4 of the Group 3 proteins and B that for the carrot II protein.

apolar and polar residues appear concentrated on a separate face of the helix, conveying an amphiphilic property to the helix; yet a small twist is given to each face resulting from the 11 residue periodicity. Figure 8 presents the same 55 residue helices in the 'helical net' diagram [4]. In A, the methyl function residues (which are the total apolar residues of this consensus sequence) are circled. In B, the methyl function residues are

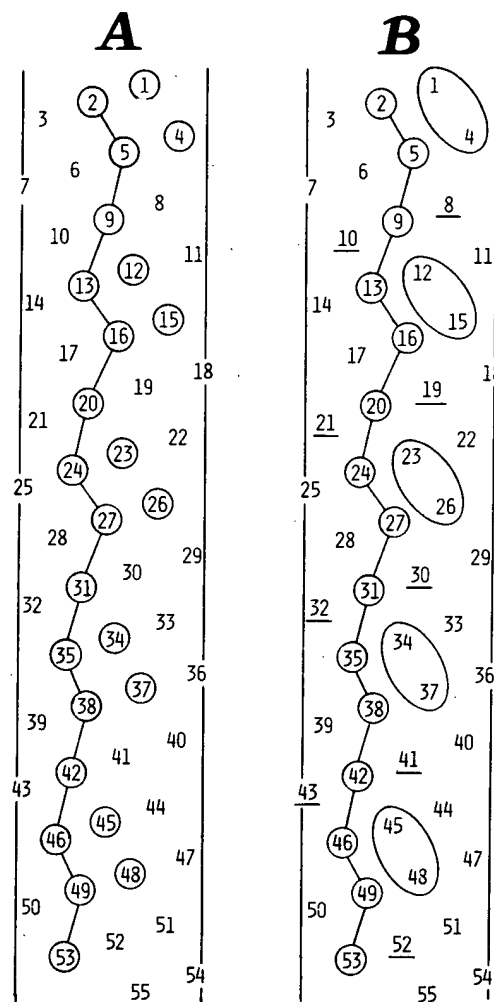


Fig. 8. Helical net diagrams of the 11-mer repeating units of the Group 3 LEA proteins. A represents the repeating units of 4 of this group and B that of the carrot II protein. The N terminus of the repeating unit is at the top. Amino acids with the methyl group function are circled. In B, other apolar amino acid functions are underlined and possible salt bridged residues joined with an ellipse. The methyl spines found in both diagrams are linked with a line.

circled and the other apolar residues are underlined. The B helix shows the 5 potential salt bridges between positive [1, 12, 23, 34 and 45] and negative [4, 15, 26, 37 and 48] residues whose formation might stabilize such a helix [11]. Salt bridges between the other oppositely charged residues of this helix, such as 6 and 7, are not likely. What is common to both helices is a slightly twisting spine of methyl groups parallel to the helix axis. Intramolecular or, more interestingly, intermolecular docking of these helices, driven by the apolar spine, to form tertiary filamentous structures is a distinct possibility.

### Discussion

The desiccation of plant seeds is characterized by the accumulation of the LEA proteins which are not found abundantly in other organs of the plant nor in embryonic tissues of seeds at earlier stages of embryogenesis. Yet, where tested, the mRNAs for many of the LEA proteins (and presumably the proteins themselves) can be induced to detectable levels in other organs or in seed tissues at other times in development by water stress and/or exposure to exogenous ABA. Desiccation stress is a normal aspect of seed formation, yet dangerous at other phases of a plant's life cycle. Thus, if LEA proteins, or a subset of them, are a part of a water stress response system, the system is a normal part of the seed developmental program, but on-call in tissues at other times in the life cycle. We suggest that such a subset of LEA proteins, once proven to function in the stress response, be named WSPs, mnemonic for water stress proteins, analogous to HSPs (heat shock proteins).

The tracts of amino acids found to be homologous among proteins of each of the 3 groups compared here may lead to the recognition of common secondary or tertiary structures as more LEA proteins are compared. Ultimately their function in protecting plant tissues in the desiccated or water-stressed state may become apparent from these structures. Towards this end, a function in coping with desiccation has been

proposed for several other cotton LEA proteins [1] for which no homologous counterparts have been found in other plants to date.

More immediately, the reality of the postulated  $\alpha$  helices of the group 3 LEA proteins based on the 11-mer repeated motif awaits demonstration by CD spectra and the X-ray crystallographic analysis of the purified proteins.

### Acknowledgements

The authors thank G. Franz for permission to use the carrot II cDNA sequence (Dc-8) before publication.

### References

1. Baker JC, Steele C, Dure III L: Sequence and characterization of 6 Lea proteins and their genes from cotton. *Plant Mol Biol* 11: 277-291 (1988).
2. Berge SK, Bartholomew RS, Quatrano RS: Control of maturation genes in wheat by ABA. In: Goldberg R (ed) *Molecular Basis of Plant Development* UCLA Symposium. J Cellular Biochemistry, New Series, Alan R. Liss, New York (in press) (1989).
3. Choi JH, Liu L, Borkird C, Sung ZR: Cloning of genes developmentally regulated during plant embryogenesis. *Proc Natl Acad Sci USA* 84: 1906-1910 (1987).
4. Dunnill P.: The use of helical net-diagrams to represent protein structures. *Biophys J* 8: 865-875 (1968).
5. Dure III L, Greenway SC, Galau GA: Developmental biochemistry of cottonseed embryogenesis and germination. XIV Changing mRNA populations as shown by in vitro and in vivo protein synthesis. *Biochemistry* 20: 4162-4168 (1981).
6. Galau GA, Hughes DW, Dure III L: Absciscic acid induction of cloned cotton late embryogenesis abundant (Lea) mRNAs. *Plant Mol Biol* 7: 155-170 (1986).
7. Harada J, DeLisle A, Baden C, Crouch M: Unusual sequence of an absciscic acid-inducible mRNA which accumulates late in *Brassica napus* development. *Plant Mol Biol* 12: 395-401 (1989).
8. Hong B, Uknes SJ, Ho T-H D: Cloning and characterization of a cDNA encoding a mRNA rapidly induced by ABA in barley aleurone layers. *Plant Mol Biol* 11: 495-506 (1988).
9. Kyte J, Doolittle RF: A simple method for displaying the hydropathic character of a protein. *J Mol Biol* 157: 105-132 (1982).
10. Litts JC, Colwell GW, Chakerian RL, Quatrano RS: The nucleotide sequence of a cDNA clone encoding the

- wheat Em protein. Nucl Acids Res 15: 3607-3618 (1987).
11. Marqusee S, Baldwin RL: Helix stabilization by Glu - ... Lys + salt bridges in short peptides of de novo design. Proc Natl Acad Sci USA 84: 8898-8902 (1987).
  12. Mundy J, Chua N-H: Absciscic acid and water stress induce the expression of a novel rice gene. EMBO J 7: 2279-2286 (1988).
  13. Roseman MA: Hydrophilicity of polar amino acid side-chains is markedly reduced by flanking peptide bonds. Mol Biol 200: 513-522 (1988).
  14. Schiffer M, Edmundson AB: Use of helical wheels to represent the structure of proteins and to identify segments with helical potential. Biophys J 7: 121-135 (1967).

tuations in initial  $^{234}\text{U}/^{238}\text{U}$  ratio is probably thousands of years), we are unable to take advantage of replicate analyses to diminish the  $^{234}\text{U}$  age uncertainties. Nonetheless,  $^{234}\text{U}$  ages are more precise than the corresponding  $^{230}\text{Th}$  ages for samples older than 450 ka and provide a useful check on the validity of the oldest  $^{230}\text{Th}$  ages. The agreement between  $^{234}\text{U}$  ages and  $^{230}\text{Th}$  ages (Table 1) for the samples that formed >350 ka (and especially for those that formed >500 ka), independently confirms the  $^{230}\text{Th}$  ages, and also shows that the growth rate of the vein for its first 100,000 years was similar to its long-term average growth rate.

Overall, the U-series ages form a remarkably self-consistent suite of age determinations. Because this consistency is both internal (from replicate samples) and external (from the stability of the overall age-distance trend), it seems highly unlikely that the dates have been significantly corrupted by open-system processes such as uranium gain or loss or alpha-recoil phenomena. The apparent ideality of the U-Th system in the vein material is probably the result of continuous submergence in water that showed limited secular variation of its physical and chemical properties (2).

## REFERENCES AND NOTES

1. J. J. Winograd *et al.*, *Science* **258**, 255 (1992).
2. A. C. Riggs, in preparation.
3. J. H. Chen, R. L. Edwards, G. J. Wasserburg, *Earth Planet. Sci. Lett.* **80**, 241 (1986); R. L. Edwards, J. H. Chen, G. J. Wasserburg, *ibid.* **81**, 175 (1986).
4. Isotopic analyses were done with a VG-54E single-collector mass spectrometer using an analogue-mode Daly detector, controlled by the program ANALYST [K. R. Ludwig, *U.S. Geol. Surv. Open-File Rep.* 85-141 (1985)]. U was analyzed in automatic mode with  $^{235}\text{U}^+$  beams of  $\sim 5 \times 10^{-14}$  A ( $^{238}\text{U}/^{235}\text{U}$  ratio assumed to be 137.88); Th was typically analyzed with  $^{230}\text{Th}^+$  beams of  $\sim 10^{-15}$  A.
5. Internal variance is variance in isotopic ratios from within-run, MS measurement only. External variance is variance in isotopic ratios not accounted for by the internal variance.
6. Amplifier nonlinearity was calibrated by measurement of fractionation-corrected  $^{235}\text{U}/^{238}\text{U}$  ratio over an ion-beam range of  $5 \times 10^{-16}$  to  $4 \times 10^{-13}$  A. A slightly nonlinear response of the Daly detector was detected, with  $\sim 0.3\%$  less gain at  $4 \times 10^{-13}$  A than at  $5 \times 10^{-16}$  A. This (software-corrected) change in gain was approximately linear with beam size, and corrections were much less than measurement precision. A regression of  $^{234}\text{U}/^{235}\text{U}$  ratio versus ion-beam intensity for hundreds of blocks of standard runs shows no resolvable residual nonlinearity. Mass fractionation during measurements was corrected by normalizing to measured  $^{233}\text{U}/^{238}\text{U}$  ratios (from the 1:1  $^{233}\text{U}/^{238}\text{U}$  spike). No fractionation correction was applied to  $^{230}\text{Th}/^{229}\text{Th}$  or  $^{232}\text{Th}/^{229}\text{Th}$  ratios, because the average effect on  $^{230}\text{Th}/^{229}\text{Th}$  was canceled out by the similarly uncorrected spike calibrations, and fractionation corrections on the  $^{230}\text{Th}/^{232}\text{Th}$  ratio were unimportant. Each Re filament was doped with graphite before use, heated to running temperature in the mass spectrometer, and examined for mass-spectral purity. Isobaric interferences were absent during data acquisition, and blank levels of  $^{230}\text{Th}$  were close to zero (average =  $-0.8 \pm 4.0 \times 10^6$  atoms). Blanks for  $^{238}\text{U}$  were negligible (median =  $1.2 \times 10^{-11}$  g), although blanks for  $^{232}\text{Th}$  were significant (median =  $4.6 \times 10^{-12}$  g). Tails under the small peaks from nearby large peaks were insignificant (abundance sensitivity at 1 atomic mass unit offset was 2 to 4 ppm), as were amplifier time-constant effects.

7. Uraninite standard HU-1, obtained from M. Ivanovich, yielded U-Pb isotopic ages of 620 million years ago, concordant to within 0.5% (K. R. Ludwig, unpublished data). Alpha-spectrometric measurements yielded  $^{234}\text{U}/^{238}\text{U}$  activity ratios of  $0.9982 \pm 0.0088$  and  $^{230}\text{Th}/^{234}\text{U}$  activity ratios within 0.5% of unity (J. N. Rosholt, personal communication to M. Ivanovich, 1979). The Precambrian age, concordance of U/Pb isotopic ages, and alpha-spectrometric measurements all indicate that the HU-1 standard is in, or very close to, secular equilibrium for  $^{230}\text{Th}/^{238}\text{U}$  and  $^{234}\text{U}/^{238}\text{U}$  ratios.
8. MSWD (mean square of weighted deviates) = 0.54.
9. MSWD = 2.0.
10. For a  $^{238}\text{U}$  half-life of  $4.4683 \times 10^9$  years: Jaffey, K. R. Flynn, L. E. Glendenin, W. C. A. M. Essling, *Phys. Rev. C* **4**, 1889 (1971).
11. P. de Bièvre *et al.*, in *Proceedings of the International Conference on Chemical and Nuclear Measurement and Applications*, M. L. H. (Institute of Civil Engineers, London, 1979).

## Claim 67 ADP-glucose pyrophosphorylase

# Regulation of the Amount of Starch in Plant Tissues by ADP Glucose Pyrophosphorylase

David M. Stark, Kurt P. Timmerman,\* Gerard F. Barry, Jack Preiss, Ganesh M. Kishore†

Starch, a major storage metabolite in plants, positively affects the agricultural yield of a number of crops. Its biosynthetic reactions use adenosine diphosphate glucose (ADPGlc) as a substrate; ADPGlc pyrophosphorylase, the enzyme involved in ADPGlc formation, is regulated by allosteric effectors. Evidence that this plastidial enzyme catalyzes a rate-limiting reaction in starch biosynthesis was derived by expression in plants of a gene that encodes a regulatory variant of this enzyme. Allosteric regulation was demonstrated to be the major physiological mechanism that controls starch biosynthesis. Thus, plant and bacterial systems for starch and glycogen biosynthesis are similar and distinct from yeast and mammalian systems, wherein glycogen synthase has been demonstrated to be the rate-limiting regulatory step.

The  $\alpha$ -1,4 glucans (starch and glycogen) are the main storage carbohydrates in practically all living systems (1). In several crops, starch is a major component of the harvest and thus directly has an impact on yield. Within the last 10 years, the demand for starch has dramatically increased for both specialized food and industrial uses (2), primarily as a result of the development of high fructose corn syrups and bio-ethanol. A number of specialty starches (such as amylose and waxy starch) are being

increasingly recognized for their superior material and nutritional properties as well as biodegradability. Understanding the critical components of the plant starch biosynthetic machinery therefore has a major impact on agriculture and industry. We have used transgenic plants to probe the rate-limiting step in starch biosynthesis.

Starch biosynthesis occurs in the plastids of plant cells, involving ADPGlc pyrophosphorylase (E.C. 2.7.7.27), starch synthase (E.C. 2.4.1.21), and branching enzyme (E.C. 2.4.1.18) (1, 3). In view of its sensitivity to allosteric effectors, ADPGlc pyrophosphorylase (ADPGPP) has been suggested to play a pivotal role in plant starch biosynthesis, as it is in the bacterial pathway for glycogen biosynthesis.

The *Escherichia coli* ADPGPP, encoded by the *glgC* gene (4), is a regulated ho-

D. M. Stark, K. P. Timmerman, G. F. Barry, G. M. Kishore, Monsanto Company, 700 Chesterfield Village Parkway, St. Louis, MO 63198.  
J. Preiss, Department of Biochemistry, Michigan State University, East Lansing, MI 48824.

\*Present address: Pioneer Hi-Bred, 7300 Northwest 62 Avenue, Johnston, IA 50131.

†To whom correspondence should be addressed.



motetrameric enzyme that is activated by fructose 1,6-bisphosphate (FbP) (5). FbP increases the catalytic activity ( $V_{max}$ ), reduces the Michaelis constant ( $K_m$ ) for the substrates, and decreases the sensitivity of the enzyme to the inhibitors adenosine monophosphate (AMP) and inorganic phosphate ( $P_i$ ) (5, 6). A mutant *E. coli* K12 strain, 618, accumulates approximately 33% higher quantities of glycogen than its wild-type parent as a result of an alteration in the regulatory properties of ADPGPP (7). This mutant enzyme is less dependent on the activator, FbP, and is less sensitive to inhibition by the inhibitor AMP. The mutant *glgC16* gene encodes a protein with a substitution of aspartic acid for glycine at position 336 in the ADPGPP enzyme (8, 9).

Plant ADPGPPs are tetramers that contain two distinct subunits and are regulated by 3-phosphoglyceric acid (PGA) and  $P_i$  as positive and negative effectors, respectively (1). Although both large and small subunits (10–13) show homology to the *E. coli* ADPGPP enzyme, the individual subunits do not appear to be catalytically active as shown by the absence of significant enzyme activity of the maize and *Arabidopsis thaliana* starch-deficient (ADPGPP) mutants (1, 14). In view of the difficulties involved in coordinating the expression of two distinct genes, we used the *E. coli glgC* gene to probe the regulation of metabolic flux by ADPGPP. Furthermore, to minimize interference by the complex allosteric regulation of the wild-type *E. coli* gene, we used the mutant *glgC16* gene.

Because starch biosynthesis occurs in plastids, we targeted the *glgC16* gene product to plastids using a modified chloroplast transit peptide (CTP) derived from an *Arabidopsis* small subunit ribulose 1,5-bisphosphate carboxylase (*rbcS*) gene (15, 16). Uptake and processing of the fusion protein (CTP-ADPGPP) by chloroplasts was confirmed by incubation of radiolabeled fusion protein with chloroplast preparations from lettuce leaves (17) (Fig. 1).

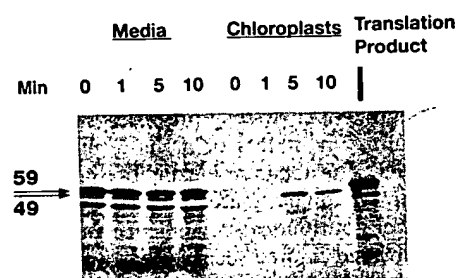
Because the *E. coli* enzyme is a homotetramer and its interaction with plant ADPGPP subunits is not known, we determined if the CTP-*glgC16* gene product is enzymatically active in plant cells using a transient expression system (18). The chimeric CTP-*glgC16* gene was introduced by electroporation into tobacco protoplasts under control of the cauliflower mosaic virus (CaMV)-enhanced 35S (e35S) promoter (19) and the polyadenylation signal derived from nopaline synthase (NOS) encoded by the nopaline synthase gene (*Nos*) (20) (Fig. 2); the activity of the gene product was assayed. Extracts of protoplasts electroporated with the chimeric gene displayed  $P_i$ -resistant ADPGPP activity, whereas the controls had little or no activity under these

conditions (Fig. 3). Thus, the *E. coli* mutant gene could be expressed in plant cells, and the gene product retains its enzymatic activity and phosphate insensitivity.

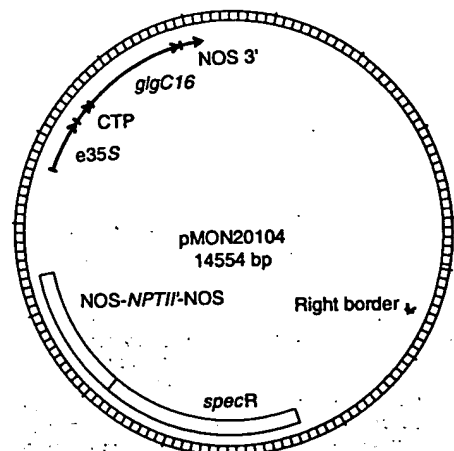
To produce stable transgenic plants that express the CTP-*glgC16* gene, we mated the pMON20104 plasmid containing a NOS-NPTII (neomycin phosphotransferase) gene as the selectable marker into *Agrobacterium tumefaciens* strain ASE and subsequently used it for the transformation of tobacco (21), tomato (22), and potato (23). Calli and shoots resulting from transformation and displaying resistance to kanamycin were analyzed for expression of the CTP-*glgC16* gene and starch content. The examination of transgenic and control tobacco calli by light microscopy revealed

substantial differences in the number of starch granules (Fig. 4). Protein immunoblot and quantitative starch analyses on 12 independent transgenic and 2 control lines showed that calli that expressed the CTP-*glgC16* gene contained on average 10.7% starch ( $SD = 6.2$ , on the basis of dry weight), a significant increase ( $P = 0.0017$ , Welch analysis of variance (ANOVA) test for unequal variance) over controls that contained 3.4% starch ( $SD = 0.5$ ). One of the samples expressing the CTP-*glgC16* gene contained 26.9% starch (Fig. 5). These observations establish that the ADPGPP reaction is the rate-limiting step in starch biosynthesis in tobacco cells. Even in tomato leaves, which typically do not accumulate large amounts of starch, a cor-

**Fig. 1.** Import and processing of the CTP-ADPGPP fusion protein by chloroplast preparations of lettuce (*Lactuca sativa* var. longifolia). Molecular size markers are indicated to the left in kilodaltons. The CTP-*glgC16* gene in pMON20100 was amplified by PCR with the use of an SP6 promoter primer (Promega, Madison, Wisconsin) and the primer used for introducing the Sac I site at the 3' end of the *glgC16* gene. The PCR product was purified by agarose gel electrophoresis and binding to a DEAE membrane. The purified DNA was incubated with SP6 RNA polymerase (Promega). The resulting RNA was translated in rabbit reticulocyte lysate (Promega) containing [ $^{35}$ S]methionine, the 19-amino acid mixture, RNA, and RNasin (Promega). SDS-gel electrophoresis of the resulting translation product demonstrated that at least two major radiolabeled proteins were produced. The slowest migrating band had a molecular size expected for the CTP-ADPGPP fusion protein (59 kD). The additional smaller bands arise from internal initiation of translation. The radiolabeled translation product was incubated for 1 to 10 min with intact chloroplast preparations from lettuce, obtained as described (17). The medium and chloroplasts in the incubation mix were separated by silicone-oil gradient centrifugation, and the chloroplast fraction was treated with trypsin. Both fractions were then treated with SDS buffer for electrophoresis. SDS-polyacrylamide gel electrophoresis was conducted with the use of a 3 to 17% polyacrylamide gradient. The gel was fixed, soaked in ENHANCE (NEN, Boston, Massachusetts), dried, and imaged by autoradiography with the use of an intensifying screen and overnight exposure at  $-70^{\circ}\text{C}$ . The 1-, 5-, and 10-min radiolabel bands in the chloroplast fractions were trypsin-resistant, which indicates that they were present within the organelle.



**Fig. 2.** A schematic representation of the chimeric 35S promoter-CTP-*glgC16*-*Nos* gene and the transferred DNA vector (pMON20104). To express the CTP-*glgC16* gene in both transient as well as stably transformed plant cells, we created a Bgl II site 7 bp upstream of the ATG translation initiation codon of the CTP in pMON20100. Xba I and Sac I sites were introduced after the termination codon in *glgC16* in the same construct. These mutations were introduced by PCR mutagenesis with the use of appropriate primers. The PCR product was digested with Bgl II and Sac I. The vector, pMON999, which contains the e35S promoter and the *Nos* 3' polyadenylation signal, was digested with Bgl II and Sac I. The product of the PCR reaction was ligated with the vector, and the result was transformed into *E. coli* MM294. The resulting plasmid, pMON20102, contained a 2.5-kb Not I fragment that consisted of the e35S promoter, CTP-*glgC16*, and the *Nos* terminator. The Not I cassette was transferred to the plant transformation vector pMON530 (34), and the resulting plasmid was designated pMON20104. The spectinomycin resistance gene is indicated by *specR*.



relation was observed between the presence of the bacterial ADPGPP and starch content (Fig. 4C).

Our attempts to regenerate plants that expressed the CTP-*glgC16* gene under the control of the *e35S* promoter resulted in poor recovery of plants expressing the gene. Our interpretation of these observations is that constitutive expression of the CTP-*glgC16* gene, as directed by the CaMV 35S promoter, is detrimental to plant growth and development. Excess starch accumulation in the leaves may reduce sucrose availability for export and thus restrict the amount of carbon available to the actively growing portions of the plant. This decrease in available sucrose also may affect the turgor of a number of different plant cells and tissues. In accord with these observations, a transgenic potato plant that expressed the *e35S*/CTP-*glgC16*/Nos gene was recovered that could be maintained on a sucrose-containing medium but could not survive in soil when sucrose was not provided in the growth medium.

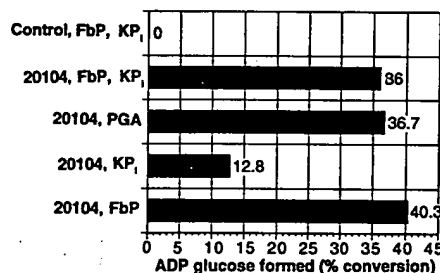
To overcome the detrimental effects of constitutive ADPGPP activity, we introduced the CTP-*glgC16* gene into potato plants under the control of a tuber-specific patatin promoter (24). We recovered numerous transgenic plants (25) with a transformation efficiency typical for potato, and plants that expressed this gene had a normal phenotype. Specific gravity measurements, which indicate tuber density and thus starch content (26), show that tubers expressing the patatin-CTP-*glgC16*

gene contain, on average, 35% more starch than control tubers (Fig. 6 and Table 1). Some transgenic potato lines produced tubers with nearly 60% more starch than controls. This range of specific gravity values in control and transgenic lines is typical of potato and has been documented (27). Chemical determination of starch content from a sampling of tubers confirmed that specific gravity was a valid reflection of starch content (Table 2). Therefore, tissue-specific expression of the CTP-*glgC16* gene with the use of a patatin promoter results in significant increases in starch content without adverse effects on plant growth or development. Recently, similar results were obtained in a small-scale field trial (25), which indicates that ADPGPP activity is rate-limiting for tuber starch biosynthesis even under conditions of potato cultivation.

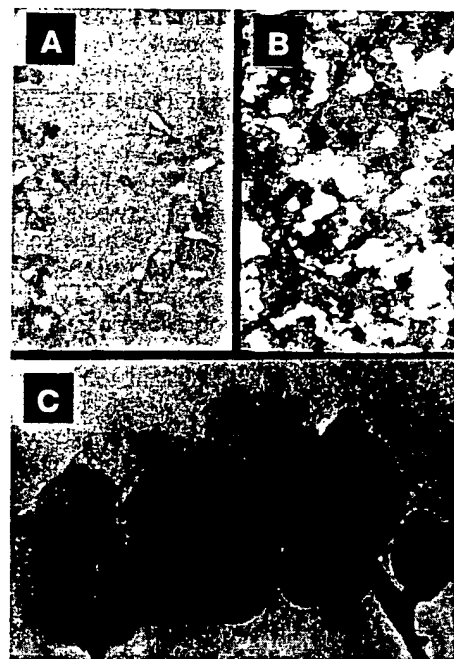
Whereas the presence of the CTP-*glgC16* gene product was absolutely required for increased starch production, the extent of starch content increase was not absolutely correlated with the level of expression of *E. coli* ADPGPP enzyme protein in all the tubers (Table 3). This suggests that expression of small amounts of the CTP-*glgC16* gene may be sufficient to overcome the limitation of ADPGlc availability for starch biosynthesis and to create a new rate-limiting step. Alternatively, substrate availability (glucose 1-phosphate) for the ADPGPP reaction may itself become rate limiting. Which of these mechanisms is operative under physiological conditions can now be investi-

gated using these transgenic potato plants.

In vitro studies that use *Commelina* guard-cell chloroplast extracts have shown that the activity of starch synthase is only one-twentieth that of ADPGPP and the branching enzymes (1, 28). These results imply that starch synthase is the rate-limiting reaction in starch biosynthesis. In view of our finding that enhanced ADPGPP activity increases the starch content of plant cells, we reasoned that allosteric regulation may down-regulate ADPGPP activity under in vivo conditions. To evaluate the importance of allosteric regulation in the control of ADPGPP activity and starch content, we fused the wild-type *E. coli glgC*



**Fig. 3.** ADPGPP activity in protoplasts. Plasmid pMON20104 was subject to electroporation into tobacco TXD protoplasts (18). Plasmid pMON999 was used as the control. The protoplasts were separated from the medium by centrifugation and washed with 100 mM tris EDTA buffer (pH 7.5) that contained 35 mM KCl and 20% glycerol. The pellet was suspended in the extraction buffer [200  $\mu$ l of washing buffer that contained 5 mM dithiothreitol (DTT), 1 mM benzamidine, and 5 mM sodium ascorbate per 50  $\mu$ l of protoplast volume] and sonicated for 2 min. The resulting suspension was centrifuged to remove the insoluble pellet, and the supernatant was desalted on a Sephadex G-50 (Sigma) spin column equilibrated with extraction buffer. The enzyme assay mix contained in 100  $\mu$ l: 10  $\mu$ mol Hepes (pH 7.7), 50  $\mu$ g of bovine serum albumin, 0.05  $\mu$ mol  $^{14}$ C-glucose 1-phosphate, 0.15  $\mu$ mol of adenosine 5'-triphosphate, 0.5  $\mu$ mol of  $MgCl_2$ , 0.1  $\mu$ g of crystalline yeast inorganic pyrophosphatase, 1 mM ammonium molybdate, and water. Fructose 1,6-bisphosphate (FbP) (2.5 mM) or 3-phosphoglycerate (20 mM) was added as indicated to activate the *E. coli* or plant ADPGPP enzymes, respectively. To further distinguish endogenous ADPGPP activity from that of the *E. coli* ADPGPP mutant, we added 10 mM inorganic phosphate ( $KP_i$ ), which completely inhibits the plant enzyme but does not affect the *E. coli* enzyme. The enzymatic reaction was carried out at 37°C for 10 min and stopped by incubation in a boiling water bath for 1 min. After centrifugation to remove the denatured protein, 40  $\mu$ l of supernatant was loaded on a SynChropak AX100 (SynChrom, Lafayette, Indiana) column (250  $\times$  4.6 mm) and eluted with 65 mM  $KP_i$  (pH 5.5) at a flow rate of 1 ml per minute. The radioactivity of the eluate was followed with the use of a Radiomatic (Radiomatic Instruments, Tampa, Florida) detector. Glucose 1-phosphate eluted at 8.3 min, and ADPGlc eluted at 14 min under these conditions. The radioactivity in the ADPGlc peak was a measure of the ADPGPP activity of the sample. Protein content of the extract was measured by the Bradford method (35) with the use of the Bio-Rad reagent.



**Fig. 4.** (A) Polarizing photomicrograph of a cross section of tobacco callus tissue transformed with vector control DNA. Leaf disks of tobacco were transformed (21) with *Agrobacterium* strain ASE containing the pMON530 vector. After 2 to 3 weeks, calli formed, and the individual clumps were separated from the leaf disks. The starch granules appear as white, birefringent structures. (B) Polarizing photomicrograph of a cross section of tobacco callus tissue transformed as in (A) with pMON20104. (C) Iodine-stained transgenic tomato shoot expressing the CTP-*glgC16* gene (right) and a tomato shoot not containing the CTP-*glgC16* gene. pMON16927 was constructed by ligation of the CTP-*glgC16* chimeric gene from pMON20102 as a Bgl II-Sac I fragment into the binary vector pMON977 (34). Cotyledons of *Lycopersicon esculentum* cv. UC82B were transformed (22) by *Agrobacterium* strain ABI containing pMON977 for control and pMON16927 for CTP-*glgC16* expression. After 2 to 3 weeks, shoots were cleanly excised from the callus, decolorized in 70% ethanol for 1 hour, and stained in a solution of 0.2%  $I_2$  and 0.4% KI.

gene with the CTP and expressed it in potato tubers and in tomato leaves using the patatin and *Arabidopsis rbcS* promoters, respectively. Unlike the GlgC16 enzyme, wild-type GlgC is fully subject to allosteric regulation (4). However, the primary effector molecules differ between the plant and bacterial enzymes. Thus, depending on the relative concentrations of these effectors in vivo, expression of the CTP-glgC gene may also result in increased ADPGPP activity and starch content in a plant cell environment. Expression, enzyme activity, and plastid targeting from the cloned glgC gene were confirmed before plant transformation (29). A large number of tomato plants transformed with the *rbcS*-CTP-glgC gene were obtained and were phenotypically normal. Analyses of leaf tissue showed high levels (0.1% of total protein) of expression of the CTP-glgC gene but only a slight effect on starch content (Table 4). The increase in average starch content of 11.2% was not significant (*t* test,  $P = 0.3111$ ). By contrast, recovery of plants with leaf expression of CTP-glgC16 was very poor, similar to the lethal effect seen with the *e35S* promoter.

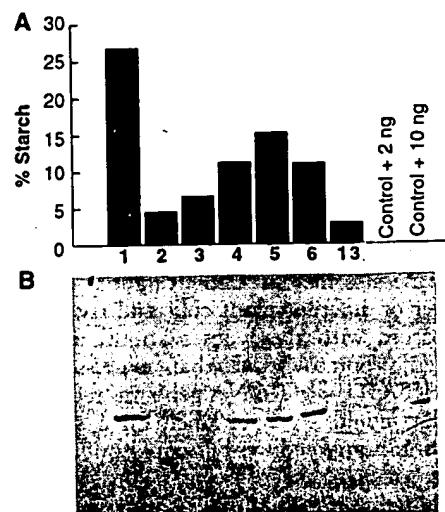
In potato tubers, expression of CTP-glgC did not result in a noticeable increase in starch content (Table 5), even though expression levels were equivalent to those in tubers that expressed CTP-glgC16. We conclude that CTP-glgC expression may lead to a slight, but not significant, amount of starch increase in plant tissues. The kinetic characteristics of the GlgC and GlgC16 enzymes have been reported (7-9) and differ primarily in allosteric regulation. The  $V_{max}$  of each enzyme is the same under fully activated conditions. On the basis of these features, our work represents further evidence that ADPGPP activity is rate limiting in starch biosynthesis and that it is the

regulatory properties of ADPGPP, not the amount of enzyme protein, that make it rate limiting.

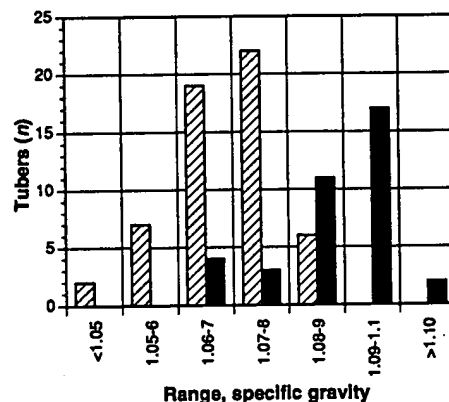
We investigated the effect of cytosol-targeted expression of GlgC16 by localizing the glgC16 gene product within the cytosol of

potato tuber cells. Amounts of glgC16 gene product were slightly smaller than amounts of the CTP-glgC16 gene product (0.02 to 0.05% versus 0.02 to 0.1% of total tuber protein) but were within the range of expression seen in high-starch CTP-glgC16 tubers. Cytosolic lo-

**Fig. 5.** Starch content (A) and protein immunoblot analysis (B) of tobacco calli. Samples 1 through 6 express the *glgC16* gene, sample 13 is a control lacking the *glgC16* gene, and the last two lanes are sample 13 with either 2 or 10 ng of GlgC16 protein purified from *E. coli*. Starch levels in transformed tobacco calli were quantitated with the method of Lin *et al.* (14). Differences are significant with the use of the Welch ANOVA *t* test for unequal variance ( $P = 0.0017$ ). For protein immunoblot analysis, a portion of the dried, homogeneous callus from each of the experimental and control samples was suspended in 200  $\mu$ l of extraction buffer [100 mM tris-Cl (pH 7.1), 1 mM EDTA, 10% glycerol, 5 mM DTT, and 1 mM benzamidine]. We ground each sample to extract the protein and centrifuged the samples to remove insoluble debris. The protein concentration of the supernatant was estimated with the use of Bradford reagent, and 25  $\mu$ g of protein from each sample was loaded on SDS-polyacrylamide gels, with a 7 to 17% gradient. After electrophoresis, the gels were blotted onto nitrocellulose, and the *E. coli* ADPGPP was visualized with the use of rabbit anti-ADPGPP as the primary antibody and goat anti-rabbit serum conjugated to alkaline phosphatase as the secondary antibody. The reaction of alkaline phosphatase with nitro blue tetrazolium and 5-bromo-4-chloro-3-indoyl phosphate was used for color development.



**Fig. 6.** Specific gravity of potato tubers expressing CTP-glgC16. We generated pMON20113 by replacing the CaMV 35S promoter in pMON999 with a 1-kb Hind III-Bam HI fragment that contained a portion of a class I patatin promoter (24). The *glgC16* gene from pMON20104, including the chloroplast transit peptide derived from the ribulose 1,5-biphosphate carboxylase small subunit 1A gene of *Arabidopsis*, was then added as a Bgl II-Sac I fragment, and the entire patatin promoter-CTP-glgC16-NOS 3' end cassette was cloned as a Not I fragment to the binary vector pMON886 (34). Potato var. Russet Burbank was transformed with *Agrobacterium* strain ASE containing the pMON20113 vector (23). Control transformations were performed with the use of the vector pMON886. Tubers from regenerated plants were screened for the expression of GlgC16 by protein immunoblot (Fig. 5). Fifteen CTP-glgC16-positive and 21 control plants were screened, and specific gravity was determined for two to three tubers per plant. Solid and hatched bars represent the numbers of tubers for those expressing *E. coli* ADPGPP and for control tubers, respectively, that fell within the indicated range of values. Controls consisted of a combination of tubers transformed to contain only the vector DNA and tubers from the *glgC16* transformation event that do not express *E. coli* ADPGPP.



**Table 1.** Average specific gravity and starch content is increased in tubers that express the CTP-glgC16 gene. The number of independent plant lines tested is indicated, with two or three tubers per plant weighed. Specific gravity was determined by the weight-in-air-weight-in-water method (26). Sample standard deviations (SD) of the specific gravity measurements are indicated. Percent starch was calculated from the average specific gravity as described (26). Controls consist of a combination of Russet Burbank tubers transformed to contain only the vector DNA and tubers from the *glgC16* transformation event that do not express *E. coli* ADPGPP.

Tuber type	Plant lines	Average specific gravity	SD	Average % starch
CTP-GlgC16	15	1.088	0.012	15.4
Control	21	1.068	0.010	11.4

**Table 2.** Average values for percent starch determined experimentally by enzymatic degradation (14) compared to starch content calculated from specific gravity measurements. Sample standard deviations are in parentheses. Differences between the methods for starch determination are not significant. However, differences between CTP-GlgC16+ and controls, calculated by specific gravity or enzymatic methods, are significant by the *t* test ( $P < 0.0001$ ). Tuber types are as in Table 1.

Tuber type	Sample size	Average % starch specific gravity	Average % starch enzymatic
CTP-GlgC16	11	16.3 (1.47)	16.0 (2.00)
Control	11	11.9 (1.56)	12.3 (1.15)

**Table 3.** Comparison of expression levels of CTP-*glgC16* and their effects on starch content. CTP-*glgC16* levels were estimated from protein immunoblot analysis by comparison to known standards. Starch content was determined by the enzymatic degradation method. One tuber [Russet Burbank (R.B.)] from each line was used for both analyses. Control starch levels represent the average value for 12 tubers.

Tuber type	Range (ng of CTP- <i>glgC16</i> per 50 µg of protein)	Range (% starch, fresh weight)
R.B. control	0	5.4 to 14.4
R.B. + high <i>GlgC16</i>	26 to 50	14.4 to 19.2
R.B. + medium <i>GlgC16</i>	10 to 25	11.8 to 19.1
R.B. + low <i>GlgC16</i>	0.5 to 10	8.8 to 17.4

**Table 4.** Starch content of tomato leaves expressing CTP-*glgC*. The wild-type CTP-*glgC* gene was engineered for expression in plants essentially as described for the CTP-*glgC16* gene (16) (Figs. 2 and 6). The CTP-*glgC* gene was added to a derivative of the binary vector pMON977 (34), which contains the *Arabidopsis* promoter from the *rbcS* gene (15) in place of e35S, resulting in the plasmid pMON16938. *Lycopersicon esculentum* cv. UC82B (L.e.) was transformed (22) with *Agrobacterium* strain ABI containing the vector pMON16938. The subsequent R1 generation was screened for the expression of CTP-*GlgC* in leaves by protein immunoblot. Standard deviations (SD) are indicated. Starch content was determined (14) in leaf tissue taken from plants at the end of the light cycle. Values are not significantly different (*t* test, *P* = 0.3111).

Leaf type	Plants ( <i>n</i> )	Average % starch (fresh weight)	SD
L.e. + CTP- <i>GlgC</i>	14	9.3	1.8
L.e. control	10	8.4	2.7

calization of the *glgC16* gene product, however, did not result in a significant effect on starch content (Table 6).

Recently, there has been controversy concerning the subcellular location of ADPGlc synthesis in plant cells. Because under in vitro conditions cytosol-localized sucrose synthase uses ADP as a substrate and plastids import ADPGlc, it has been suggested that sucrose synthase serves as a source of ADPGlc for starch biosynthesis in plant cells and tissues (30). However, in view of our demonstration that only the expression of CTP-*glgC16*, not *glgC16*, re-

sults in an increased starch content, it is unlikely that this mechanism contributes significantly to starch biosynthesis. Further evidence in support of the plastidial ADPGPP pathway comes from recent work that shows that reduction of the plant ADPGPP enzyme level by means of antisense RNA results in decreased starch formation (31).

Because potato tubers that express the CTP-*glgC16* gene contain more starch, it would be interesting to determine if this is a result of the ability of the potato plant to provide additional photosynthate to the tuber or a result of a reduction in water content of the tuber due to rapid conversion of the osmotically active sugars to starch. Evidence exists that plants contain unused photosynthetic capacity that could become available should the demand be present (32). This is supported by numerous reports that show that increased carbon demand results in increased assimilation in and output from the leaves (33).

ADPGPP has been suggested to influence the ability of tubers to import carbon and convert it into starch—that is, the rate of starch accumulation regulates the rate of carbon import (1); this hypothesis is now supported by our study. The transgenic potato plants described here will allow the testing of models that suggest that photosynthetic carbon assimilation is not limiting to yield. It is interesting that a single enzymatic step regulates the end-product levels to such a dramatic extent in a complex multicellular organism. With a similar approach, it should be possible for

**Table 6.** Specific gravity and starch content of potato tubers that express *GlgC16* in the cytosol. A cytosolic version of the *glgC16* gene was created by ligation of the coding region from pMON20100 as a Nco I and Sac I fragment into a version of pBluescript kS+ (Stratagene, La Jolla, California) engineered to contain Bgl II and Nco I sites in the polylinker. The coding region was then ligated as a Bgl II and Sac I fragment into the binary vector pMON16952, which contains the class I patatin promoter, resulting in the plasmid pMON16971. Potato var. Russet Burbank was transformed with *Agrobacterium* strain ABI containing the pMON16971 vector (23). Tubers formed from regenerated plants were analyzed as described (Table 1).

Tuber type	Plant lines	Average specific gravity	SD	Average % starch
<i>GlgC16</i> +	26	1.073	0.010	12.4
Controls	22	1.073	0.008	12.4

researchers to manipulate the amounts of a number of key metabolites such as lipids, amino acids, and carbohydrates by affecting or altering a major regulatory step in the biosynthesis of relevant metabolites.

## REFERENCES AND NOTES

1. J. Preiss, in *The Biochemistry of Plants*, J. Preiss, Ed. (Academic Press, Orlando, FL, 1988), pp. 184–249; in *Oxford Survey of Plant Molecular and Cellular Biology*, B. J. Milfin, Ed. (Oxford Univ. Press, Oxford, 1991), vol. 7, pp. 59–114.
2. F. R. Katz, in *Biotechnology and Food Ingredients*, I. Goldberg and R. Williams, Eds. (Van Nostrand Reinhold, New York, 1991), pp. 315–326.
3. E. Beck and P. Zeigler, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 40, 95 (1989).
4. T. W. Okita, R. L. Rodriguez, J. Preiss, *J. Biol. Chem.* 256, 6944 (1981).
5. J. Preiss, L. Shen, E. Greenberg, N. Genter, *Biochemistry* 5, 1833 (1966).
6. N. Genter and J. Preiss, *J. Biol. Chem.* 243, 5882 (1968); J. Preiss and T. Romeo, *Adv. Bact. Physiol.* 30, 184 (1989).
7. J. Cattaneo, M. Damotte, N. Sigal, F. Sanchez-Medina, J. Puig, *Biochem. Biophys. Res. Commun.* 34, 694 (1969); N. Creuzet-Sigal, M. Latil-Damotte, J. Puig, in *Biochemistry of the Glycosidic Linkage*, R. Piras and H. G. Pontis, Eds. (Academic Press, New York, 1972), vol. 2, pp. 647–680.
8. P. Leung et al., *J. Bacteriol.* 167, 82 (1986).
9. Y. M. Lee, A. Kumar, J. Preiss, *Nucleic Acids Res.* 15, 10603 (1987); J. Preiss and M. A. Hill, unpublished results.
10. J. M. Anderson et al., *J. Biol. Chem.* 264, 12238 (1989).
11. M. R. Olive, R. J. Ellis, W. W. Schuch, *Plant Mol. Biol.* 12, 525 (1989).
12. M. R. Bhawe, S. Lawrence, C. Barton, L. C. Hannah, *Plant Cell* 2, 581 (1990).
13. B. T. Muller, J. Koßmann, L. C. Hannah, L. Willmitzer, U. Sonnewald, *Mol. Gen. Genet.* 224, 136 (1990); P. A. Nakata et al., *Plant Mol. Biol.* 17, 1089 (1991).
14. T. P. Lin, T. Caspar, C. Somerville, J. Preiss, *Plant Physiol.* 86, 1131 (1988).
15. M. P. Timko et al., *Nature* 318, 579 (1985).
16. The *Arabidopsis* CTP was generated with the use of the chloroplast-targeting and mature NH<sub>2</sub>-terminal coding regions of the ribulose carbox-

**Table 5.** Starch content of potato tubers that express CTP-*glgC*. The CTP-*glgC* gene was ligated into a derivative of the binary vector pMON977 containing the patatin promoter in place of e35S, resulting in the plasmid pMON16950. Potato var. Russet Burbank was transformed with *Agrobacterium* strain ABI containing the pMON16950 vector (23). Analysis of tubers from regenerated plants was as described (Table 1). Starch content was determined as in Table 1.

Tuber type	Plant lines	Average specific gravity	SD	Average % starch
R.B. + CTP- <i>GlgC</i>	6	1.076	0.005	13.1
R.B. control	17	1.077	0.010	13.2

ylase 1A gene (*rbcS* 1A) (15). The CTP was modified by addition of the first 23 amino acids (MQVWPPIGKKKFFETLSYLPDLTDS) because the efficiency of protein import is increased by addition of the sequences from the mature region of this protein. Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. To facilitate removal of the remaining *rbcS* sequences from the ADPGPP protein, we added a second protease cleavage site (GGRVNMQA) between these 23 amino acids of *rbcS* and the NH<sub>2</sub>-terminal Met of *E. coli* ADPGPP. The *glgC16* gene was obtained as a Hinc II fragment from plasmid pLP226 (8) and cloned into the pUC8 vector at the Hinc II site. To fuse the modified transit peptide coding region to the translation initiation site of the *glgC16* gene, we introduced an Nco I site at the initiating methionine of the *glgC16* gene. Also, to remove 3' nucleotide sequences, we added a Sac I site downstream of the termination codon of the *glgC16* gene. Both Nco I and Sac I sites were introduced by polymerase chain reaction (PCR) mutagenesis. The modified *Arabidopsis* CTP plus the *glgC16* gene were cloned into pGEM3zf+ (Promega, Madison, WI), digested with Hind III and Sac I, by ligating the *Arabidopsis* CTP as a Hind III-Nco I fragment and *glgC16* as an Nco I-Sac I fragment. The resulting plas-

mid (pMON20100) consisted of pGEM3zf+, the modified *Arabidopsis* CTP, and the *glgC16* gene. The CTP-*glgC16* gene in the plasmid was in an orientation suitable for transcription by the SP6 promoter resident in the pGEM3zf+ plasmid.

17. G. della-Cioppa *et al.*, *Biotechnology* 5, 579 (1987).
18. J. Callis, M. Fromm, V. Walbot, *Nucleic Acids Res.* 15, 5823 (1987).
19. R. Kay, A. Chan, M. Daly, J. McPherson, *Science* 236, 1299 (1987).
20. M. Bevan, *Nucleic Acids Res.* 12, 8711 (1984).
21. R. B. Horsch *et al.*, *Science* 223, 496 (1984).
22. S. McCormick *et al.*, *Plant Cell Rep.* 5, 80 (1986).
23. C. A. Newell *et al.*, *ibid.* 10, 30 (1991).
24. M. Bevan *et al.*, *Nucleic Acids Res.* 14, 4625 (1986).
25. D. M. Stark, G. F. Barry, Y. M. Muskopf, K. P. Timmerman, G. M. Kishore, paper presented at the Third International Congress of Plant Molecular Biology, Tucson, AZ, 6–11 October 1991, abstract 714.
26. C. von Scheele, *Landwirtsch. Vers. Stn.* 127, 67 (1937).
27. G. E. Kleinkopf, D. T. Westermann, M. J. Wille, G. D. Kleinschmidt, *Am. Potato J.* 64, 579 (1987).
28. N. L. Robinson and J. Preiss, *Plant Physiol.* 85, 360 (1987).
29. D. M. Stark, K. P. Timmerman, G. F. Barry, J.

Preiss, G. M. Kishore, data not shown.

30. J. Pozueta-Romero, M. Frehner, A. M. Viale, T. Akazawa, *Proc. Natl. Acad. Sci. U.S.A.* 88, 5769 (1991).
31. B. Muller-Rober, U. Sonnewald, L. Willmitzer, *EMBO J.* 11, 1229 (1992).
32. R. B. Dwell, *Am. Potato J.* 67, 829 (1990).
33. W. S. Cranshaw and E. B. Radcliffe, *J. Econ. Entomol.* 73, 131 (1980); R. B. Dwell, in *Potato Physiology*, P. Li, Ed. (Academic Press, Orlando, FL, 1985), pp. 35–58; E. E. Ewing, K. P. Sandlan, A. G. Nicholson, paper presented at the 11th Triennial Conference of the European Association for Potato Research, Edinburgh, Scotland, 1990; J. Moorby, *Symp. Soc. Exp. Biol.* 31, 425 (1977).
34. S. G. Rogers, H. J. Klee, R. B. Horsch, R. T. Fraley, in *Methods in Enzymology*, R. Wu and L. Grossman, Eds. (Academic Press, San Diego, 1987), pp. 253–277.
35. M. Bradford, *Anal. Biochem.* 72, 248 (1976).
36. We thank N. Mathis, Y. Muskopf, J. Heinz, and J. Layton for plant transformation; J. Hart, B. LaValle, B. Schuler, T. Coombe, M. Pratt, and B. Coombs for maintenance of plants; M. Hayford for pMON999; M. Bevan for the patatin promoter; C. Hironaka for the CTP; M. Hinchey for assistance in light microscopy; and R. T. Fraley for support.

29 April 1992; accepted 31 July 1992

## Intercellular Propagation of Calcium Waves Mediated by Inositol Trisphosphate

Scott Boitano, Ellen R. Dirksen, Michael J. Sanderson\*

Two types of calcium ( $\text{Ca}^{2+}$ ) signaling—propagating intercellular  $\text{Ca}^{2+}$  waves of increasing intracellular  $\text{Ca}^{2+}$  concentration ( $[\text{Ca}^{2+}]_i$ ) and nonpropagating oscillations in  $[\text{Ca}^{2+}]_i$ —co-exist in a variety of cell types. To investigate this difference in  $\text{Ca}^{2+}$  signaling, airway epithelial cells were loaded with heparin, an inositol 1,4,5-trisphosphate ( $\text{IP}_3$ ) receptor antagonist, by pulsed, high-frequency electroporation. Heparin inhibited propagation of intercellular  $\text{Ca}^{2+}$  waves but not oscillations of  $[\text{Ca}^{2+}]_i$ . In heparin-free cells,  $\text{Ca}^{2+}$  waves propagated through cells displaying  $[\text{Ca}^{2+}]_i$  oscillations. Depletion of intracellular  $\text{Ca}^{2+}$  pools with the  $\text{Ca}^{2+}$ -pump inhibitor thapsigargin also inhibited the propagation of  $\text{Ca}^{2+}$  waves. These studies demonstrate that the release of  $\text{Ca}^{2+}$  by  $\text{IP}_3$  is necessary for the propagation of intercellular  $\text{Ca}^{2+}$  waves and suggest that  $\text{IP}_3$  moves through gap junctions to communicate intercellular  $\text{Ca}^{2+}$  waves.

Intercellular communication is essential for the function of multicellular systems, but the nature of the signal or signals that pass between cells through gap junctions is not fully established. Both  $\text{Ca}^{2+}$  and  $\text{IP}_3$  have been proposed as intercellular messengers (1–3). Nonexcitable cells often respond to agonists by increasing their  $[\text{Ca}^{2+}]_i$  in an oscillatory manner (4), but these oscillations in  $[\text{Ca}^{2+}]_i$  occur independently of  $[\text{Ca}^{2+}]_i$  changes in adjacent cells (3). In contrast to  $\text{Ca}^{2+}$  oscillations, a propagating intercellular wave of increased  $[\text{Ca}^{2+}]_i$  (a  $\text{Ca}^{2+}$  wave) can be initiated by mechanical stimulation of a single cell in cultures of airway epithelial (2), rat brain glial (3, 5,

6), or bovine aortic endothelial (7) cells or by treatment of astrocytes with glutamate (8). In airway epithelial cells  $\text{Ca}^{2+}$  waves are blocked by the gap junction inhibitor  $\text{Ca}^{2+}$  halothane (2), and in C6 glioma cells only cells transfected with and expressing the gene for the gap junction protein connexin43 propagate  $\text{Ca}^{2+}$  waves (6). These results indicate that  $\text{Ca}^{2+}$  waves are propagated through gap junctions (1–3, 5–8). A role for  $\text{IP}_3$  in the communication of  $\text{Ca}^{2+}$  waves has been proposed because  $\text{Ca}^{2+}$  waves are propagated in the absence of extracellular  $\text{Ca}^{2+}$  (2, 3, 7), are propagated when  $\text{Ca}^{2+}$ -induced  $\text{Ca}^{2+}$  release is inhibited (5), and can be initiated by microinjection of  $\text{IP}_3$  (2). If  $\text{IP}_3$  acts as the intercellular messenger for propagation of  $\text{Ca}^{2+}$  waves, intracellular heparin, an antagonist of the  $\text{IP}_3$  receptor (9), should block or

attenuate the  $\text{Ca}^{2+}$  wave (9).

Traditional loading techniques, such as microinjection, are not well suited for loading the large numbers of cells required for investigation of this multicellular response. Therefore, we used pulsed high-frequency electroporation (PHFE) (10, 11) to load cultured cells with heparin (Fig. 1). Because heparin cannot be detected by fluorescence microscopy, cells were simultaneously loaded with fluorescent Texas red-conjugated dextran (TRD) to identify cells that incorporated heparin (Fig. 1) (11, 12). PHFE has a major advantage over current-discharge electroporation in that most cells survive (10). After PHFE, and loading cells with fura-2 by incubation in fura-2-pentaacetoxymethyl ester (fura-2-AM) (12), more than 90% of the heparin-loaded cells retained fura-2 and heparin-loaded ciliated cells continued to display ciliary activity. These results demonstrate the potential of PHFE for loading cells with molecules that are impermeable to the cell membrane because of their large molecular size or ionic charge.

In an area where all cells were loaded with heparin, mechanical stimulation of a single cell increased  $[\text{Ca}^{2+}]_i$  in the stimulated cell but did not initiate propagation of a  $\text{Ca}^{2+}$  wave through multiple adjacent cells, even though the increase in  $[\text{Ca}^{2+}]_i$  of the stimulated cell ranged from 300 nM to  $>1 \mu\text{M}$  (Fig. 2A). In a few of these experiments in the heparin-loaded area,  $[\text{Ca}^{2+}]_i$  increased in single cells directly adjoining the stimulated cell, but an increase of  $[\text{Ca}^{2+}]_i$  in more distal cells was not observed. The increase in  $[\text{Ca}^{2+}]_i$  of the stimulated, heparin-loaded cells and our reports that the  $[\text{Ca}^{2+}]_i$  of a mechanically

Department of Anatomy and Cell Biology, UCLA School of Medicine, Los Angeles, CA 90024.

\*To whom correspondence should be addressed.

# Cloning of the yeast *FAS3* gene and primary structure of yeast acetyl-CoA carboxylase

(acetyl-CoA carboxylase gene/biotin carboxylase/transcarboxylase)

WALID AL-FEEL, SUBRAHMANYAM S. CHIRALA, AND SALIH J. WAKIL\*

Verna and Marrs McLean Department of Biochemistry, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030

Contributed by Salih J. Wakil, December 24, 1991

**ABSTRACT** We have isolated and determined the nucleotide sequence of the yeast *FAS3* gene, which encodes acetyl-CoA carboxylase (EC 6.4.1.2). The sequence has an open reading frame of 6711 bases coding for a protein of 2237 amino acids with a calculated molecular weight of 250,593. The presence of the unique biotin-binding site, Met-Lys-Met, and the known CNBr peptide and COOH-terminal sequences confirmed the nucleotide-derived amino acid sequence. The yeast, chicken, and rat carboxylases have an overall sequence identity of 34%, suggesting that the eukaryotic carboxylase evolved from a single ancestral gene. The amino acid sequences of yeast fatty acid synthase subunits are least homologous with the animal synthase sequences, whereas carboxylase sequences are highly conserved. The sequences of the ATP,  $\text{HCO}_3^-$ , and CoA binding sites of the carboxylases are also well conserved ( $\approx 50\%$  identical). The sequences surrounding the biotin binding site are poorly conserved, suggesting that this sequence may not be critical as long as the biotin is available for carboxylase reactions. On the basis of this sequence identity, we have defined the putative biotin carboxylase and transcarboxylase domains.

Acetyl-CoA carboxylase (ACC; EC 6.4.1.2) catalyzes the committed step in fatty acid biosynthesis, yielding malonyl-CoA, the donor of the two-carbon units for the synthesis of long-chain fatty acids (1). In prokaryotes, the enzyme consists of three readily dissociated proteins, the biotin carboxyl carrier protein (BCCP), the biotin carboxylase, and the transcarboxylase (2), whereas in higher and lower eukaryotic cells, these proteins are part of a single multifunctional polypeptide derived from the expression of a single gene that, presumably, evolved by the fusion of individual genes. Recently, the cDNAs coding for the rat (3) and chicken (4) ACC were cloned and sequenced. The sites for the biotin attachments in both carboxylases, which are conserved as in all other biotin-containing enzymes, were readily identified. Putative domains for the two catalytic functions were assigned (3, 4) and sites for phosphorylation were located (5).

ACC has been isolated from *Saccharomyces cerevisiae* (6) and *Candida lipolytica* (7) and shown to contain single-subunit proteins of molecular weights 189,000 and 230,000, respectively. To understand the relationship between structure and function of the yeast ACC and to utilize the potential of genetic manipulations in yeast, we have undertaken a systematic analysis of the enzyme. In this report, we describe the isolation and nucleotide sequence of the *FAS3* gene,<sup>†</sup> which encodes yeast ACC, and compare the deduced amino acid sequence with sequences of the rat and chicken enzymes to determine their evolutionary relationship.

## MATERIALS AND METHODS

**Preparation of ACC and Its Antibodies.** Yeast ACC from extracts of baker's yeast was purified to a state of homogeneity. Cell-free extracts were prepared, as described previously (8), and the enzyme was isolated by ammonium sulfate fractionation (0–28% saturation), polyethylene glycol 8000 (0–6%) precipitation, and avidin-Sepharose affinity chromatography (9). The carboxylase preparation had specific activity of 2.5 units/mg protein when assayed by [ $^{14}\text{C}$ ]bicarbonate incorporation into malonyl-CoA, as described (9). Antibodies against the purified ACC were raised in rabbits, and affinity-purified anti-ACC antibodies were prepared (10).

**Isolation and Sequencing of the *FAS3* Gene.** The yeast genomic DNA libraries in  $\lambda\text{gt}11$  and EMBL3a vectors were provided by M. Snyder (Yale University) (11). The  $\lambda\text{gt}11$  library was screened with anti-ACC antibodies by following standard procedures (11, 12), and the EMBL3a library was screened with radioactive DNA probes as described (13). The yeast strain SEY2102 was grown in appropriate media and total RNA was isolated for Northern analysis (14).

DNA sequencing was performed by using both the dideoxynucleotide termination method (15), as described previously (14), and an automated DNA sequencer (Applied Biosystems model 370A), according to manufacturer's recommendations. All the restriction enzymes and other chemicals were purchased from commercial and standard sources (9, 14).

## RESULTS AND DISCUSSION

**Isolation and Expression of  $\lambda\text{gt}11\text{acc}$ .** The native yeast ACC is a tetramer of identical subunits each having an estimated molecular weight of 250,000 (Fig. 1). Affinity-purified anti-ACC antibodies were used to screen a yeast  $\lambda\text{gt}11$  genomic DNA expression library (11, 12). A positive clone,  $\lambda\text{gt}11\text{acc}$ , was isolated and shown to contain a 3.0-kilobase-pair (kbp) fragment of the putative ACC genomic DNA. Initial verification came from DNA sequence analysis of an *Sst* I–*Sst* I fragment from  $\lambda\text{gt}11\text{acc}$  (see Fig. 3). The nucleotide-derived amino acid sequence showed a high degree of sequence identity with animal ACC DNA sequences, and, on the basis of this homology, it was concluded that the  $\lambda\text{gt}11\text{acc}$  clone contained the portion of the *FAS3* gene coding for the COOH terminus of yeast ACC. The identity of this clone was verified further by immunoblotting (16) of protein lysates from  $\lambda\text{gt}11\text{acc}/\text{Y1089}$  lysogens (10). As shown in Fig. 1, a fusion protein ( $M_r \approx 180,000$ ) produced by the  $\lambda\text{gt}11\text{acc}$  recombinant phage reacted with the anti-ACC antibodies. On the basis of the sizes of the fusion protein and the  $\beta$ -galactosidase, the carboxylase gene fragment in  $\lambda\text{gt}11\text{acc}$  coded for

Abbreviations: ACC, acetyl-CoA carboxylase; BCCP, biotin carboxyl carrier protein.

\*To whom reprints should be addressed.

<sup>†</sup>The sequence reported in this paper has been deposited in the GenBank data base (accession no. M92156).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.





ATGACGGAAGAAAGCTTATTCGAGTCTTCCACAGAAGATGGAGTACGAAATACAACTACTCAGAAAGACATACAGAACTCCAGGTCATTTTCATTCGGCTC  
MetSerGluGluSerLeuPheGluSerSerProGlnLysMetGluTyrGluIleThr~~Asn~~TyrSerGluArgHisThrGluLeuProGlyHisPheIleGlyLeu 35

AATACAGTAGATAAATACAGGAGTCCCGCTTAAGGAGCTTTGTTAAGAGTACCGGTGCTCACACGGTCATATCCAGATCTGATAGCAAAATAAGTATTCGC  
AsnThrValAspLysLeuGluSerGlnLysLeuArgAspPheValLysSerHisGlyGlyHisThrValIleSerLysIleLeuIleAlaAsnAsnGlyIleAla 70

GCCGTGAAGAAATAGATCCCTCAGAAAATGGGCATACGAGCGTTCGGCGATGACAGAACCGTCCAATTCGTCGCCATGGCCACCCAGAAATCTCGAGCGCC  
AlaValLysGluIleArgSerValArgLysTrpAlaTyrGluThrPheGlyAspAspArgThrValGlnPheValAlaMetAlaThrProGluAspLeuGluAla 105

AACGCGAATATATCCGTATGGCCGATCAATACATTGAAGTCCGAGGTGGTAACTAATAACAACTACGCTAACGCTAGACTTGATCGTAGACATCGCCGAAAGA  
AsnAlaGluTyrIleArgMetAlaAspGlnTyrIleGluValProGlyGlyThrAsnAsnAsnAsnTyrAlaAsnValAspLeuIleValAspIleAlaGluArg 140

GCAGCGTAGACCGCTATGGCGTGGTGGGTCACGCTCCGAGAAATCCACTATTCCTGAAAAATTCCTCCAGTCTAAGAGGAAAGTCATCTTTATTCGGCCT  
AlaAspValAspAlaValTrpAlaGlyTrpGlyHisAlaSerGluAsnProLeuLeuProGluLysLeuSerGlnSerLysArgLysValIlePheIleGlyPro 175

CCAGCTAACGCCATGAGGTCTTAGGTGATAAAATCTCTCTACCATTCGCTCAAGTCTCAAGTCCCATGTTATCCATGCTCTGCTACCGGTGTTGACACC  
ProGlyAsnAlaMetArgSerLeuSerGlnLysIleSerSerThrIleValAlaGlnSerAlaLysValProCysIleProTrpSerGlyThrGlyValAspThr 210

GTTACGTCGACGAGAAAACCGGTCTGGTCTCTGTCGACGATGACATCTATCAAAACGGTGTGTCTACCTCTCTCAAGATGCTTTTCAAAAAGGCCAAGCGTATT  
ValHisValAspGluLysThrGlyLeuValSerValAspAspAspIleTyrGlnLysGlyCysCysThrSerProGluAspGlyLeuGlnLysAlaLysArgIle 245

GCTTTTCCTGTCATGATTAAGCGATCCGAAGGTGGTGGTAAAGGTATCAGACAAGTTCAGACGTAAGAAAGATTTCATCCGCTTTATACCAACCGCAGCCAAAC  
GlyThrProValThrIleArgMetAlaGlyGlyGlyGlyIleArgGlnValGluArgGluGluAspPheIleAlaLeuTyrHisGlnAlaAlaAsn 280

GAAATTCAGGCTCCCCATTTTCATCATGAAGTTGGCGGTAGAGCGGCTCACTTGGAAAGTTCACCTGACGATCAGTACCGTACAAATATTTCTCTGTTT  
GluIleProGlySerProIlePheIleMetLysLeuAlaGlyArgAlaArgHisLeuGluValGlnLeuLeuAlaAspGlnTyrGlyThr~~Asn~~IleSerLeuPhe 315

GGTAGAGACTCTTCGTCAGACGCTCATCAAAAAATTTATCGAAGAAGCACCAGTTACAATTTGCCAAGGCTGAAACATTTACAGAGATGGAAAAGGCTCGCCCT  
GlyArgAspCysSerIleLeuSerGluArgHisGlyGlyIleGluGluAlaProValThrIleAlaLysAlaGluThrPheHisGluMetGluLysAlaAlaVal 350

AGACTGGGAAATAGTCGGTTATGCTCTGCGGTACCGTGGAGTATCTATATTTCTCATGATGATGGAAAAATTTACTTTTTAGAAATGAACCAAGATTACAA  
ArgLeuGlyLysLeuValGlyTyrValSerAlaGlyThrValGluTyrLeuTyrSerHisAspAspGlyLysPheTyrPheLeuGluLeuAsnProArgLeuGln 385

GTCGAGCATCAACACCGAAATGCTCTCGGTGTTAACTACCTGCGCTCAATTTACAAATCGCTATCGGTATCCCTATGCATAGAATAAGTGACATTAGAAT  
ValGluHisProThrThrSerGluAspProAsnAspGlyPheLysProSerGlyGlyThrLeuHisGluLeuAsnPheArgSerSerSerAsnValTrpGlyTyrPhe 420

TTATATGGTATGAATCTCTCTGCTCAGAAATCGATTTCGAATTCAAAATCAAGATGCCCAAGAAACAAAGAACCTATTCCAAAGGCTCATTTGATACC  
LeuTyrGlyMetAsnProHisSerAlaSerGluIleAspPheGluPheLysThrGlnAspAlaThrLysLysGlnArgArgProIleProLysGlyHisCysThr 455

GCTTCTGTCATCATCAGAGATCCAAACGATGGATTCAAGCCATCGGGTGGTACTTTCGCACTGAACTAAATTCGCTTCTCTCTTAATGTTTGGGTTACTTCT  
AlaCysArgHisPheThrIleSerGluSerGluAsnGlyGlyPheLysProSerGlyGlyThrLeuHisGluLeuAsnPheArgSerSerSerAsnValTrpGlyTyrPhe 490

TCCGTGGTAACAAATGGTAATTTCACTCTCTTCGGACTCTCAGTTCGGCCATATTTTGTCTTTGCTGAAATAGACAAGCTTCCAGGAAACACATCGTTGTT  
SerValGlyAsnAsnGlyAsnIleHisSerPheSerAspSerGlnPheGlyHisIlePheAlaPheGlyGluAsnArgGlnAlaSerArgLysHisMetValVal 525

GCCCTGAAGGAATTTGCCATTAGGGGTGATTTCAGAACTACTGTGAAATCTTGATCAAACTTTTGGAACTGAAGATTTCCAGGATAACACTATTACCCCGGT  
AlaLeuHisProLeuSerIleArgGlyAspArgThrThrValGluTyrLeuIleLysLeuLeuGluThrGluAspPheGluAspAsnThrIleThrThrGly 560

TGGTGGACGATTGATTACTCATAAATGACCGCTGAAAGCCTGATCAACACTCTTCCCGTCAATTCGCGTCCCGCTACAAAGCCTTCTTAGCATCTGAAGAA  
TrpLeuAspAspLeuIleThrHisLysMetThrAlaGluLysProAspProThrLeuAlaValIleCysGlyAlaAlaThrLysAlaPheLeuAlaSerGluGlu 595

GCCCGCCACAAGTATATCGAATCTTTACAAAAGGCAAGCTTCTATCAAGACCTACTGCAAACTATGTTCTCTCTAGATTATTTATCATGACGGTAAAGATAC  
AlaArgHisLysTyrIleGluSerGluGlnLysGlyGlnValLeuSerLysAspLeuLeuGlnThrMetPheProValAspPheIleHisGluGlyLysArgTyr 630

AAGTTACCGTAGCTAAATCCGGTAATGACCGTTACACATTATTTATCAATGGTTCTAAATGTGATATCATCTCGCTCAACTATCTGATGGTGGTCTTTTGATT  
LysPheThrValAlaLysSerGlyAsnAspArgTyrThrLeuPheIle~~Asn~~GlySerLysCysAspIleIleLeuArgGlnLeuSerAspGlyGlyLeuLeuIle 665

GCCATAGGCGGTAAATCGCATACCATCTATTTGAAAGAAGAGTTCCTCTACAGAGTATTCGGTCTACTATGACTTCTTGTGCAAGTTGAAAACGATCCA  
AlaIleGlyGlyLysSerHisThrIleTyrTrpLysGluValAlaAlaThrArgLeuSerValAspSerMetThrThrLeuLeuGluValGluAsnAspPro 700

ACCCAGTTGCGTACTCCATCCCTGGTAAATGGTTAAATCTTGGTGGAAAAATGGTGAACACATATCAAGGCCCAACCATATGCAGAAATGAAGTTATGAAA  
ThrGlnLeuArgThrProSerProGlyLysLeuValLysPheLeuValGluAsnGlyGluHisIleIleLysGlyGlnProTyrAlaGluIleGluValMetLys 735

ATGCAATGCTCTTGGTTTCTCAAGAAATGGTATCGTCCAGTTATTAAGCAACCTGGTCTTCACTTGTTCAGGTGATATCATGGCTATTTAGCTCTTGAC  
MetGluHisProLeuLeuLysGlyTyrAspAsnGlyIleValGlnLeuLysGlnValIleMet~~Asn~~AlaSerLeuGlnGlnLeuIleGluValLeuArgAsnProLysLeuProTyr 770

GATCCATCCAAGGTCAAGCAGCTCTACCATTTGAAGGTATGCTGCCAGATTTTGGTCTCCAGTTATCGAAGCAACCAACCTGCCTATAAATTCAGTCATTA  
AspProSerLysValLysHisAlaLeuProPheGluGlyMetLeuProAspPheGlySerProValIleGluGlyThrLysProAlaTyrLysPheLysSerLeu 805

GTGTCTACTTTGGAACAACTTTGAAGGGTTATGACAAACCAAGTTATTATGAAACGCTTCTTCAACAAATGATAGAGTTTTCAGAAATCCAAAATCGCTTAC  
ValSerThrLeuLysHisIleLeuLysLeuArgAspGluAsnProLysAspAsnGlnValIleMet~~Asn~~AlaSerLeuGlnGlnLeuIleGluValLeuArgAsnProLysLeuProTyr 840

TCAGAAATGGAACATCACATCTCTGCTTTACATTCAAGATTGCTGCTAAGCTAGATGAACAAATGGAAGAGTTAGTTGCACGTTCTTTGAGACGCTGCTGCTT  
SerGluTrpLysLeuHisIleSerAlaLeuHisSerArgLeuProAlaLysLeuAspGluGlnMetGluGluLeuValAlaArgSerLeuArgArgGlyAlaVal 875

TTCCAGCTAGACAAATTAAGTAAATTTGATGATGCGCGTGAAGAAATCCTGAATACAAACCCGACAAATTCGTCGGCGCGCTGCGAACCAATTCGCGGATATT  
PheProThrLeuLysHisIleLeuLysLeuArgAspMetAlaValLysAsnProGluTyrAsnProAspLysLeuLeuGlyAlaValValGluProLeuAlaAspIle 910

GCTCATAAGTACTCTAAGCGGTAGAAGCCCATGAACATCTTATATTTGTCATTTCTTGAAGAAATATTACGAAGTTGAAAAGTTATTCAATGGTCCAAATGTT  
AlaHisLysTyrSerAsnGlyLeuGluAlaHisGluHisSerIlePheValHisPheLeuGluGluTyrTyrGluValGluLysLeuPheAsnGlyProAsnVal 945

CGTGAGAAAAATATCTTGAAATTCGGGTGATGAAGAACCTAAGATCTAGATAAAGTTGCGCTAACTGTTTCTCTCATTGGAAGTTTCAGCGAAGAAATAAC  
ArgGluGluAsnIleIleLeuLysLeuArgAspGluAsnProLysAspLeuAspLysValAlaLeuThrValLeuSerHisSerLysValSerAlaLysAsnAsn 980

CTGATCTAGCTATCTTGAACATATCAACCATTTGTGCAAGTTATCTTCTAAGTTTCTGCTTCTTCTACTCTCTACCAACATATGTTGAACTAGAACTCT  
LeuIleLeuAlaIleLeuLysHisTyrGlnProLeuLysLysLeuSerSerLysValSerAlaIlePheSerThrProLeuGlnHisIleValGluLeuGluSer 1015

AAGCTACCGCTAAGGTGCTCTCAACGAGAGAAATTTGATTCAAGCGCTTTACCTTCGCTCAAGGAAAGAACTGAACAAATGAACATATCTTAAATCC  
LysAlaThrAlaLysValAlaLeuGlnAlaArgIleLeuIleGlnGlyAlaLeuProSerValLysGluArgThrGluGlnIleGluHisIleLeuLysSer 1050

TCTGTGTGAAGGTGCTTATGCTCATCAATCCAAAGCGCTCTGAACAGATTGAAATATCTTGAAGACTTGATCGATTCTAATACGTTGCTGTTGATGTT  
SerValValLysValAlaTyrGlySerSerAsnProLysArgSerGluProAspLeuAsnIleLeuLysAspLeuIleAspSerAsnTyrValValPheAspVal 1085

TTACTCAATTTCTAACCCATCAAGACCCAGTTGTGACTGCTGACGCTGCTCAAGTCTATATTCGTCGCTTATCGTCTTACCACTAGAGATATTAGAGTT  
LeuLeuGlnPheLeuThrHisGlnAspProValThrAlaAlaAlaGlnValTyrIleArgArgAlaTyrArgAlaTyrThrIleGlyAspIleArgVal 1120

CACGAAGGTGCACAGTTCCAATTTGTAATGGAATTTCAACTACCTTCAGTCTGCTTCTCCACCTTTCCAACCTGTTAAATCTAAAAATGGTATGAACAGGGCT  
HisGluGlyValThrValProIleValGluTyrLysPheGlnLeuProSerAlaAlaPheSerThrPheProThrValLysSerLysMetGlyMetAsnArgAla 1155

GTTTCTGTTTCAGATTGTCATATGTTGCAACAGTCAGTATCTCGTTAAGACAGGATTTTGTATGCGTGTGATCATTTAGATGATGTTGATGAAATTTTG  
ValSerValSerAspLeuSerTyrValAlaAsnSerGlnSerSerProLeuArgGluGlyIleLeuMetAlaValAspHisLeuAspAspValAspGluIleLeu 1190

TCACAAAGTTTGAAGTTATTTCTGTCACCAATCTTCTTAACGCGACTGCTCTGATCGTTCTGAGTCTCGCATCGTTGAGTAAATGTTGCTAATGTTTGT  
SerGlnSerLeuGluValIleProArgHisGlnSerSerSerAsnGlyProAlaProAspArgSerGlySerSerAlaSerLeuSerAsnValAlaAsnValCys 1225

GTTGCTTCTACAGAAAGTTTCAATCTGAAGAGGAAATTTGGTAAGTTGAGAGAAATTTGGATTGAAATGAAGAGAAATTAATCAATGCTTCTATCCGTCGT  
ValAlaSerThrGluGlyPheGluSerGluGluGluIleLeuValArgLeuArgGluIleLeuAspLeuAsnLysGlnGluLeuIle~~Asn~~AlaSerIleArgArg 1260

ATCACATTTATGTTCCGTTTAAAGATGGGTCTTATCCAAAGATTATATCTTTAAACGGTCCAAATTAACGAAATGAACAAATTCGTCACATTGAGCCGGCT  
IleThrPheMetPheGlyPheLysAspGlySerTyrProLysTyrTyrThrPheAsnGlyProAsnTyrAsnGluAsnGluThrIleArgHisIleGluProAla 1295

TTCGCTTCCAACTGGAATAGGAAGATTGTCACCACTCAACATTAACCAATTTTCACTGATAATAGAAATCCATGCTACGAAGCTGTTAGTAAGACTTCT  
LeuAlaPheGlnLeuGluGlyArgLeuSerAsnPheAsnIleLysProIlePheThrAspAsnArgAsnIleHisValTyrGluAlaValSerLysThrSer 1330

Biotin  
CarboxylaseBiotin  
Binding  
Site

FIG. 4. (Figure continues on the opposite page.)



Trans-  
Carboxylase

FIG. 4. Nucleotide sequence of the gene coding for ACC and the predicted amino acid sequence of the protein. Numbering of the nucleotides starts with the A of the first ATG Met codon. Numbering is shown only for the amino acids. Underlined amino acid sequences indicate protein sequences that have been verified by amino acid sequencing and the glycosylation sites. The amino acid sequences in bold letters indicate the putative nucleotide-binding motif, biotin-binding site, and the putative CoA-binding site.

The polypeptide segments of the yeast ACC that exhibit high amino acid sequence similarity to those of rat enzyme vary in length and occur in three subdomains along the protein (Fig. 5). The first subdomain, which spans the region near the NH<sub>2</sub> terminus (residues 100–600) and exhibits about 50% identity to corresponding sequences of rat enzyme, could be the putative biotin carboxylase component of yeast ACC because of its high sequence similarity to corresponding amino acid sequences of known biotin enzymes such as yeast pyruvate carboxylase (residues 162–355), and the  $\alpha$  subunit of human propionyl-CoA carboxylase (residues 177–375). Moreover, within the yeast ACC subdomain, the region between amino acid residues 235 and 392 is highly homologous to a corresponding segment of rat ACC and may contain the ATP and HCO<sub>3</sub><sup>-</sup> binding sites. This conclusion was based, in part, on the presence of the Gly-rich motif of Gly-Xaa-Gly-Xaa-Xaa-Gly or Gly-Xaa-Xaa-Gly-Xaa-Gly, which have been suggested as the nucleotide-binding motif for yeast and rat carbamoyl-phosphate synthases and rat ACC (3, 22). Further, Hamada *et al.* (23) have proposed that the sequence Tyr-Gly-Tyr-Thr-His-Leu-Ser-Thr-Gly in rabbit muscle myokinase (residues 32–40) is involved in the binding of MgATP. Similar sequences are found within this subdomain

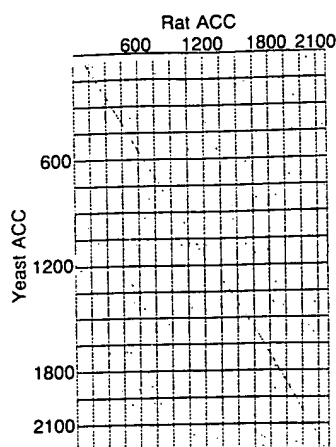


FIG. 5. Dot-matrix plot of rat vs. yeast amino acid sequences of ACC.

of yeast ACC (residues 366–373), Tyr-Leu-Tyr-Ser-His-Asp-Asp-Gly, and in rat ACC (residues 424–432), Tyr-Leu-Tyr-Ser-Gln-Asp-Asp-Ser; however, their significance in nucleotide binding remains to be determined.

The conserved biotin-binding site Met-Lys-Met is located at residues 734–736 in yeast ACC, which represents the region of biotin carboxyl carrier protein, the second subdomain. Comparison of amino acid sequences near the biotin-binding sites with all known biotin-containing enzymes showed that (i) the Met-Lys-Met sequences of yeast and animal ACCs are preceded by Val instead of Ala, as in other carboxylases, and (ii) these sequences are located closer to the NH<sub>2</sub> termini of the molecules, whereas in other biotin-containing enzymes this sequence occurs near the COOH termini. It has been suggested that this positioning of the biotin-binding site may increase the efficiency of the biotinylation of the proteins (24). In all the biotin-containing enzymes, the biocytin residue is located 25–29 amino acids downstream from a short amino acid sequence flanked by two Pro residues [-Pro-(Xaa)<sub>n</sub>-Pro-], which might act as a hinge to permit the biotin-containing arm to move between the carboxyl donor and acceptor sites (2). Yeast, rat, and chicken ACCs also contain similar sequences (Fig. 4).

Despite the conservation of biotin-binding sites -Met-Lys-Met- among all biotin-containing proteins, the amino acid sequences surrounding these sites are divergent. Indeed, the sequence of yeast ACC between residues 600 and 1700 is least homologous (26%) with that of the rat ACC (Fig. 5). The lack of conservation of the sequences in these regions of the enzymes may pertain to the assembly of the enzyme subunits into polymer forms that make up the enzymatically active carboxylases. Also, this lack of conservation suggests that the amino acid sequences within BCCP domains of the carboxylases may be involved only in providing a scaffold for the critical regions of the structure to function. In this regard, the biotin in the BCCP domain may be akin to the 4'-phosphopantetheine prosthetic group of the acyl carrier protein domain of the fatty acid synthase, where the quaternary structure of the protein and the need for the pantetheine-bound fatty acyl intermediates to interact with the various catalytic domains of the synthase play a crucial role in the overall activity of the enzymes.

The third subdomain of the yeast ACC, residues 1700–2100, is highly homologous (60%) to the corresponding segment of the rat ACC, residues 1650–2200 (3). Within these subdomains, sequences between amino acid residues 1870 and 1890 are highly similar to the proposed sequence of the "adenine recognition loop" of porcine and yeast citrate

synthase (25, 26) and to the  $\beta$  subunit of human propionyl-CoA carboxylase (27). Hence, these peptide segments in ACC are likely to be components of the CoA-binding site.

Recently, Bowers and Allred (28) reported that the rat liver ACC is a glycoprotein. In the yeast carboxylase sequence, there are nine sites that can be N-glycosylated, as indicated in Fig. 4. However, the presence of carbohydrates in the carboxylase and the involvement of any of the putative glycosylation sites remain to be investigated.

We thank C. Green, S. Richards, and K. C. Janssen for their technical help, B. Lee Ligon for editorial assistance, and W.-Y. Huang for providing the amino acid sequence of CNBr peptide. This work was supported in part by Grant DMB 8713051 from the National Science Foundation and a grant from the Clayton Foundation for Research to S.J.W. and Grant DK41872 from the National Institutes of Health to S.S.C.

- Wakil, S. J., Stoops, J. K. & Joshi, V. C. (1983) *Annu. Rev. Biochem.* **52**, 537–579.
- Wood, H. G. & Barden, R. E. (1977) *Annu. Review. Biochem.* **46**, 385–413.
- Lopez-Casillas, F., Bai, D.-H., Luo, X., Kong, I.-S., Hermodson, M. S. & Kim, K.-H. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 5784–5788.
- Takai, T., Yokoyama, C., Wada, K. & Tanabe, T. (1988) *J. Biol. Chem.* **263**, 2651–2657.
- Davies, S. P., Sim, A. T. R. & Hardie, G. H. (1990) *Eur. J. Biochem.* **198**, 183–190.
- Sumper, M. & Riepertinger, C. (1972) *Eur. J. Biochem.* **29**, 237–248.
- Mishina, M., Kamiryo, T., Tanaka, A., Fukui, S. & Numa, S. (1976) *Eur. J. Biochem.* **71**, 295–300.
- Stoops, J. K., Awad, E. S., Arslanian, M. J., Gunsberg, S., Wakil, S. J. & Oliver, R. M. (1978) *J. Biol. Chem.* **253**, 4464–4475.
- Thampy, K. G. & Wakil, S. J. (1985) *J. Biol. Chem.* **260**, 6318–6323.
- Shapiro, D. J., Taylor, J. M., McKnight, G. S., Palacios, R., Gonzales, C., Kiely, M. L. & Schimke, R. T. (1974) *J. Biol. Chem.* **249**, 3665–3671.
- Snyder, M., Elledge, S., Sweetser, D., Young, R. A. & Davis, R. W. (1987) *Methods Enzymol.* **154**, 107–128.
- Young, R. A. & Davis, R. W. (1984) *Science* **222**, 778–780.
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
- Chirala, S. S., Kuziora, M. A., Spector, D. M. & Wakil, S. J. (1987) *J. Biol. Chem.* **262**, 4231–4240.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
- Towbin, H., Staehelin, T. & Gordon, J. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 4350–4354.
- Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 5201–5205.
- Mohamed, A. H., Chirala, S. S., Mody, N. H., Huang, W.-Y. & Wakil, S. J. (1988) *J. Biol. Chem.* **263**, 12315–12325.
- Kodzak, M. (1981) *Nucleic Acids. Res.* **9**, 5233–5252.
- Lynen, F. (1979) *CRC Crit. Rev. Biochem.* **7**, 103–119.
- Kong, I.-S., Lopez-Casillas, F. & Kim, K.-H. (1990) *J. Biol. Chem.* **265**, 13695–13701.
- Fry, D. C., Kuby, S. A. & Mildvan, A. S. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 907–911.
- Hamada, M., Palmieri, R. H., Russell, G. A. & Kuby, S. A. (1979) *Arch. Biochem. Biophys.* **195**, 155–177.
- Bai, D.-H., Moon, T.-W., Lopez-Casillas, F., Andrews, P. C. & Kim, K.-H. (1989) *Eur. J. Biochem.* **182**, 239–245.
- Suissa, M., Suda, K. & Schatz, G. (1984) *EMBO J.* **3**, 1773–1781.
- Remington, S., Wiegand, G. & Huber, R. (1982) *J. Mol. Biol.* **158**, 111–152.
- Lamhonwah, A.-M., Barankiewicz, T. J., Willard, J. F., Mahuran, D. J., Quan, F. & Gravel, R. Y. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 8049–8053.
- Bowers, D. F. & Allred, J. B. (1991) *FASEB J.* **5**, A426 (abstr.).

## Molecular cloning, characterization, and elicitation of acetyl-CoA carboxylase from alfalfa

BASIL S. SHORROSH\*<sup>†‡</sup>, RICHARD A. DIXON<sup>†</sup>, AND JOHN B. OHLROGGE\*

\*Department of Botany and Plant Pathology, Plant Biology Building, Michigan State University, East Lansing, MI 48824-1312; and <sup>†</sup>Plant Biology Division, The Samuel Roberts Noble Foundation, P.O. Box 2180, Ardmore, OK 73402

Communicated by Paul K. Stumpf, December 20, 1993 (received for review October 13, 1993)

**ABSTRACT** Acetyl-CoA carboxylase [ACCase; acetyl-CoA:carbon-dioxide ligase (ADP-forming), EC 6.4.1.2] catalyzes the ATP-dependent carboxylation of acetyl CoA to produce malonyl CoA. In plants, malonyl CoA is needed for plastid localized fatty acid biosynthesis and for a variety of pathways in the cytoplasm including flavonoid biosynthesis. We have determined the full nucleotide sequence of an ACCase from alfalfa, which appears to represent a cytoplasmic isozyme. Partial cDNAs were isolated from a cDNA library of suspension culture cells that had been elicited for isoflavonoid phytoalexin synthesis. The full-length sequence was obtained by primer extension and amplification of the cDNA with synthetic primers. The sequence codes for a protein of 2257 amino acids with a calculated  $M_r$  of 252,039. The biotin carboxylase, biotin carboxyl carrier protein, and carboxyltransferase domains, respectively, show approximately 72%, 50%, and 65% sequence similarity to those of animal, diatom, and yeast ACCase sequences. ACCase enzyme activity and transcripts are induced severalfold upon addition of yeast or fungal elicitors to alfalfa cell cultures.

Acetyl-CoA carboxylase [ACCase; acetyl-CoA:carbon-dioxide ligase (ADP-forming), EC 6.4.1.2], a biotin-containing enzyme, catalyzes the first step in fatty acid biosynthesis via ATP-dependent carboxylation of acetyl CoA to form malonyl CoA. This reaction has been implicated as a key regulatory step in *de novo* fatty acid biosynthesis in the plastids of plant leaves (1–3) and in the cytosol of animal cells (4).

A single ACCase polypeptide has been purified from several higher plants, algae, and animals. However, two ACCase isozymes differing in their sensitivity toward the herbicides haloxyfop and sethoxydim have recently been identified in maize (5). In addition to providing malonyl CoA for plastid fatty acid synthesis, additional ACCase isozymes may be needed in the cytosol to supply malonyl CoA for the synthesis of very long chain fatty acids (6), flavonoids (7, 8), isoflavonoid phytoalexins (9), and stilbenes (10), and in the inactivation of the ethylene precursor 1-aminocyclopropane 1-carboxylate (11). ACCase activity is induced in UV-illuminated parsley cultures (7) and elicitor-treated soybean cultures (8) at the onset of accumulation of flavonoids and isoflavonoid phytoalexins, respectively. The increased ACCase activity in these cultures is presumed due to a higher demand for malonyl CoA for the chalcone synthase reaction.

ACCase has been cloned from *Escherichia coli* (12–15), chicken (16), rat (17), yeast (18), and the diatom *Cyclotella cryptica* (19). Recently, two subunits of *Anabaena* sp. ACCase have been cloned (20). The full-length sequence of a higher plant ACCase has not previously been reported. The *E. coli* ACCase is composed of four separate proteins: the biotin carboxyl carrier protein (BCCP), the biotin carboxylase (BC), and two subunits of carboxyltransferase (CT). In

contrast, mammalian and yeast ACCases contain the BCCP, BC, and CT as functional domains within a single polypeptide with an apparent mass of 200–250 kDa. In plants, both multisubunit (21–23) and multifunctional polypeptide forms of ACCase (3) have been reported.

We report here the isolation of alfalfa ACCase cDNAs covering the complete ACCase open reading frame, which exhibits 60–63% sequence similarity and 40–42% sequence identity at the amino acid level to chicken, rat, yeast, and *C. cryptica* ACCases. We also show that treatment of alfalfa cell suspension cultures with fungal or yeast elicitors results in substantial induction of ACCase activity and transcript levels. Based on the absence of a transit peptide in the sequence and the induction of expression by elicitors, we suggest that the alfalfa ACCase clone codes for a cytoplasmic isozyme.<sup>§</sup>

### MATERIALS AND METHODS

**Growth and Elicitation of Alfalfa Cell Cultures.** Cell suspension cultures of alfalfa (*Medicago sativa* L.) cv. Apollo were initiated and maintained as described elsewhere (24). Dark-grown cells were treated at 5 days after subculture with elicitor preparations from the cell walls of *Colletotrichum lindemuthianum* (24) or yeast (25) at final concentrations of 50- $\mu$ g glucose equivalents per ml of medium.

**Extraction and Assay of ACCase Activity.** Alfalfa cell suspension cultures (1 g) were homogenized with a Polytron on ice in 2 ml of 100 mM Mes, pH 6.5/2 mM dithiothreitol. The supernatant was brought to 45% saturation with solid  $(\text{NH}_4)_2\text{SO}_4$  and left on ice overnight. The precipitated protein was resuspended in 400  $\mu$ l of extraction buffer and 30  $\mu$ l of the extract was used immediately for ACCase assay (based on the formation of acid-stable radioactive malonyl CoA from  $\text{H}^{14}\text{CO}_3$  and acetyl CoA) as described elsewhere (21). Assays lacking acetyl CoA were included for each cell extract as a control. Protein content was determined in duplicate by the Bradford assay (26) using bovine serum albumin as a standard.

**Construction and Screening of an Alfalfa Genomic Library.** Genomic DNA was isolated from alfalfa leaves (cv. Apollo), partially digested with *Bam*HI, filled in with dGTP and dATP, and ligated to Lambda FIX II/*Xho* I filled in with dATP and dGTP as described by the supplier (Stratagene). The ligated product was packaged with a Gigapack II Gold kit and amplified in the host strain PLK17 (Stratagene). The library was plated with the host P2PLK17 and screened with the 3'ACC probe by plaque hybridization at 42°C as described (27). Positive plaques were amplified in PLK17 and purified as described elsewhere (28). A region of the genomic

Abbreviations: ACCase, acetyl CoA carboxylase; BCCP, biotin carboxyl carrier protein; BC, biotin carboxylase; CT, carboxyltransferase.

<sup>†</sup>Present address: Department of Botany and Plant Pathology, Michigan State University, East Lansing, MI 48824-1312.

<sup>§</sup>The sequence reported in this paper has been deposited in the GenBank data base (accession no. L25042).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

DNA insert was amplified by PCR using universal primer (CAU)<sub>4</sub>T3, which anneals to the Lambda FIX II arm, and primer 180 (described below). The amplified genomic DNA was annealed to the pAMP1 vector and sequenced as described below.

**Screening of an Alfalfa cDNA Library.** A cDNA library was constructed in Lambda ZAP II starting with poly(A) RNA isolated from alfalfa cell suspension cultures treated with fungal elicitor for 2, 3, and 4 h. A cDNA fragment (585 bp) encoding a putative alfalfa ACCase was identified by its sequence (see *Results*). This fragment was used to screen the library as described (27). To find additional cDNAs overlapping the 5' end of characterized ACCase cDNAs, probes corresponding to 200–300 bp at the 5' ends of the identified cDNAs were generated by PCR (27) and used to further screen the cDNA library.

**5' Rapid Amplification of cDNA Ends (RACE) PCR for Isolation of Sequences Corresponding to the 5' End of the ACCase Transcript.** *First-strand cDNA synthesis by primer extension.* Total RNA (10 µg) was isolated from cell suspension cultures treated with yeast elicitor for 5 h and mixed with 200 ng of 30-mer primer corresponding to the 5' end of clone T1 (Fig. 1). The RNA/primer mixture was denatured at 65°C, annealed at 42°C, and reverse-transcribed in a total vol of 20 µl using M-MLV (BRL) at 42°C in the presence of 10% dimethyl sulfoxide. The RNA was hydrolyzed with 2 µl of 2 M NaOH at 65°C for 1 h and then neutralized with 2 µl of 2 M HCl.

**Poly(A) tailing and second-strand cDNA synthesis.** Four of the first-strand cDNA synthesis reaction mixtures were pooled and the cDNA was purified into 50  $\mu$ l of H<sub>2</sub>O using a GeneClean II kit (Bio 101). The cDNA (50  $\mu$ l) was tailed at its 3' end with poly(A) in a final vol of 100  $\mu$ l using 75 units of terminal deoxynucleotidyltransferase (BRL) at 37°C. The second strand was synthesized with 2.5 units of *Taq* DNA polymerase in a total vol of 100  $\mu$ l containing 50  $\mu$ l of poly(A)-tailed cDNA, 200  $\mu$ M each dNTP, 1  $\mu$ M anchor primer 134 (AAGCTTCTGCAGGAGCTCTTTTTTTTTT-TTTTTTTT), and 3 mM MgCl<sub>2</sub>. The mixture was incubated at 95°C for 3 min, 25°C for 5 min, 30°C for 10 min, 40°C for 15 min, and 72°C for 30 min.

**Nested primers used for amplification.** The second-strand cDNA reaction mixture (10  $\mu$ l) was amplified at 50°C annealing temperature with 20-mer primers that included at their 5' end the base sequence CAUCAUCAU or CUACUACUACU to facilitate cloning into the pAMP1 vector. The first PCR product was generated by primers 135 (based on primer 134 above) and 136 (corresponding to the 5' end of clone T1). The second amplification was with primers 141 (corresponding to the 5' end of the cDNA amplified with primers 135 and 136) and 142 (designed to encode the peptide

EGGGGKG with codon usage based on the earlier characterized alfalfa ACCase cDNA sequences). The third amplification was with primers 146 (corresponding to the 5' end of the cDNA amplified with primers 141 and 142) and 147 (designed to encode the peptide KVLIANNG). The peptides EGGGGKG and KVLIANNG are conserved between rat, chicken, and yeast ACCases. The fourth amplification used primer 209 (based on genomic ACCase sequences) and 180 (corresponding to the 5' end of the cDNA amplified with primers 146 and 147) (Fig. 1).

**Isolation and Sequencing of DNA Templates.** Plasmids isolated from the alfalfa cDNA library were subjected to double-strand sequencing of both strands using dideoxynucleotide chain termination (29) with T7 DNA polymerase (United States Biochemical). Each PCR-generated clone was isolated from three independent colonies and was fully sequenced on both strands using universal and synthetic primers (30) and *Taq* DNA polymerase with an ABI robotic catalyst and model 373A DNA sequencer. In some cases, the PCR product was directly subjected to sequence analysis. The alignment was determined by using the GAP program of Genetics Computer Group (University of Wisconsin, Madison).

**Northern and Southern Blot Analysis.** Total RNA (15  $\mu$ g) was isolated from alfalfa cell suspension cultures (31), separated on 1% agarose containing 2.2 M formaldehyde, and blotted onto nitrocellulose membranes. The blot was hybridized at 42°C in the presence of 50% formamide as described elsewhere (32). RNA blots were tested for equal loading by probing with alfalfa 25S rRNA or G1 cDNA probes (32, 33) (data not shown). High molecular weight genomic DNA was isolated from plant leaves (34), digested with restriction endonucleases, electrophoresed in 0.7% agarose, and blotted onto a GeneScreenPlus nylon membrane with 0.4 M NaOH. Blots were hybridized and washed at 65°C or 55°C as described elsewhere (32).

## RESULTS

**Isolation of Alfalfa ACCase cDNA Clones.** An alfalfa cDNA clone designated G1 (32), isolated by screening an alfalfa cDNA library with a human protein disulfide isomerase probe, was found to be fused through its poly(A) tail to an unrelated 585-bp cDNA fragment. This fragment (3' ACC) (Fig. 1) encoded a polypeptide with 40% amino acid sequence identity and 77% sequence similarity to the chicken ACCase in their overlapping regions. Rescreening of the alfalfa cDNA library (see *Materials and Methods*) led to the isolation of three clones: Q7 (2567 bp), M2 (4200 bp), and T1 (1467 bp) (Fig. 1). When aligned, the three clones covered 5.3 kb of the ACCase coding region and their overlapping regions were 100% identical. The cDNA library lacked clones containing the 5' coding region of the putative alfalfa ACCase. This region of the ACCase sequence was therefore obtained by the 5' RACE PCR method. This led to the isolation of four additional clones designated 135/136 (625 bp), 141/142 (835 bp), 146/147 (700 bp), and 209/180 (351 bp) (Fig. 1).

**Primary Sequence of Alfalfa ACCase.** Alignment of all the above alfalfa ACCase cDNA clones yielded a sequence of 7194 bp with 69 and 353 bp of 5' and 3' untranslated regions, respectively. The 3' untranslated region contained two potential polyadenylation signals (AATAA) starting at positions 7113 and 7124 (data not shown). Of 7 cDNA clones sequenced at the 3' untranslated ends, 1 was polyadenylated at position 7171, 4 at position 7175, and 2 after addition of the sequence TTTTAT to position 7175.

The coding sequence of 6771 bp contains a methionine residue, which initiates the longest open reading frame of 2257 amino acids encoding a protein of  $M_r$  252,039 (Fig. 2). This methionine residue has a nucleotide context similar to the Kozak initiator methionine consensus sequence (35) and

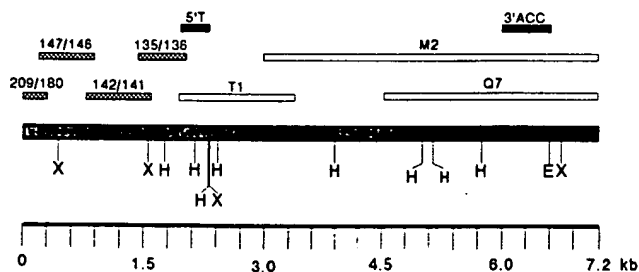


FIG. 1. Schematic representation of alfalfa ACCase sequences. Nucleotide sequence position (in kb) and partial restriction map sites are indicated. Regions used as probes for Southern and Northern blot analysis are indicated by solid bars (5'T, 3'ACC). cDNA clones are represented by open bars. PCR-generated cDNAs are represented by hatched bars. Numbers above each hatched bar represent the pair of primers used to generate the PCR product. H, *Hind*III; X, *Xba* I; E, *Eco*RV.

is preceded by a cDNA sequence rich with stop codons in all three reading frames. Based on sequence alignment with other known ACCases, the putative alfalfa ACCase protein contains three functional domains arranged, starting from the N-terminal end, in the following order: BC, BCCP, and CT (Fig. 2). A putative biotin binding site at a lysine residue flanked by two methionine residues starts at position 712 in the alfalfa ACCase protein (Fig. 2). The two proline residues present 27 and 35 residues upstream of the putative biotin

Alf	.....NASVGRGIGLNSVPSHPATTTEVDEYCNALGKPKIHSILIANHGAAYKFRISV	50
Rat	ASVSDTSLDGLSALDGLAFMWSMSGLMLVKDGRD.....RKIDSDQDFVAPAEVTRFGKVIKELYIANHGAAYKFRISV	136
Alf	RSNATYFTGTEKALVAMATPEKRIHAEIRIADQFVPGGTHNNHNYAVLLIAETIHWDAVPGHASENPELPAKAGI	148
Rat	RRVSTENFRKRIATFYVNYTPEDLKAAEYISNAHHTVPPGGANNNHNYAVLLIAETIHWDAVPGHASENPELPAKAGI	226
Alf	VFLGPAISMAALGKIGSSLIADAAYFTPLVSGSHYKIP, PESD.....ITIPDIYRAACVYTTETAIASQVYVTPNLIKASGGG	233
Rat	AFNPPSQAAMALGKIGSSLIADAAYFTPLVSGSHYKIP, PESD.....ITIPDIYRAACVYTTETAIASQVYVTPNLIKASGGG	316
Alf	GGSTGCHHDEVRALPKVGVGPGSIFINVKVASGSHVLEQLCDHGFALHSRDSVDRHKKIIEEGPIIVAPPETKLEDA	323
Rat	GGSTGCHHDEVRALPKVGVGPGSIFINVKVASGSHVLEQLCDHGFALHSRDSVDRHKKIIEEGPIIVAPPETKLEDA	406
Alf	ARLKASVHTVGAATVETLSHETGEAYFLMLRPLQVEHPVETIAELPAQVAVGNGIPLVQIPEIRKFTGHEGGGDKGKTSV	413
Rat	AVLKAKGVHTVGAATVETLSHETGEAYFLMLRPLQVEHPVETIAELPAQVAVGNGIPLVQIPEIRKFTGHEGGGDKGKTSV	486
Alf	LATPFDFDE, ASDTKHGHVAVRYTSDPGDFTGKGVKYSFKSPNNVATVSVKSGGHEFSDSGFVAFGESSRALIAADV	502
Rat	.....LATPFDFDE, ASDTKHGHVAVRYTSDPGDFTGKGVKYSFKSPNNVATVSVKSGGHEFSDSGFVAFGESSRALIAADV	574
Alf	LGLKEITIGETIRTYDITDILNADSDRDKHITGALDSRIANRYAERPNMYSVGGALYKASASSAALVSDYVGLKSGIPKHI	592
Rat	VAKKELISGDFRTIETVLEKLETESFQNRITDGLDGLAEKQVQERPTMLVGGALYKASASSAALVSDYVGLKSGIPKHI	664
Alf	SLVHSQVSLTSEGGKTYIDVYRGSGSKLKLKNSIEAEITLROGGLMLDQSHVITYAEAAAGTLLDGRGLTLDNDQDPSKI	682
Rat	LLNTVDVELTVEGTYLVKTRPSVSVYVIMHSGSCVEVDHMLSGGLLSDGSSYTYKVEEDRYITIGKCTVFEKEDSVNR	764
Alf	GETPKLRLYLVDSDQSDIADTPAEVEYVNHCMPLSPASGIIHFRMAEGAMQAGELIAKLDGGSVAKPEFTSPGPIIGPTAI	772
Rat	SPSAGLITVIEGGHFAQGTAEITVGHVNTLTAVESGCTIVKVAQALDQCEVIAHQGLDPSKQQAELHTGSLPIQISTALR	854
Alf	SGKYNKCAASLAARHILAGY.....EHRIDEVY.....VKSLLNCLSDPELFLQVCEFAVLAIRLPCDLRKLKAEKFEIISSTQID	866
Rat	GEKLRHYFTVLDNLVYHNGYCLPOFFSKVSKVQVRLKRLDPSLLELQDINTSVSGRIPLNVEKISKEKAGYASHTVSLQD	934
Alf	FRALKKALLENHLS, CPENEGKALRELVPELTVKSYEGGEGSHAKIVQSLVEEYLSVEELSDHITQADVIELRLDLYKCLKIV	945
Rat	FPQSQIATLDSMAATLNRKEREVFWNTQSIQVQVRSGLRSHAKVQVRLDLYKVEYVQNGHYKQVCFALKEKESQNTV	1024
Alf	DIVLSHQVSKKLLILRLKMLVTPNPA.....AYRQLIRFSQNLHIVSEALAKASQLLEQTLK.....ELRSSIARS.....LSELNFTDQ	1028
Rat	NYTFSHAGVYKLLVTLTDLQCGDPLTDELLILTELTLQSKITRAVLAARQVLTASHLPSYDVRQVQVESITLSAIDKTHQ	1113
Alf	ENIDTPKRSATDRHEDVSLAPLADALVGLFQNSDHTLQRRVYETIRLTPYTKDSIRHMRSGLIATMEFLEETVEKRGV	1118
Rat	.....FCIEKQLLSESTIFDVLPHFFYSKQVRYVMAALVYVRYATYATLNSVQRLKONTCYVEFQKPLTSHPNRGL	1193
Alf	.....DKTLEKSE.....KKGWVYVYKSLQFLPATISALREATHKPFKPSGSSGSHGAMHILGLV	1182
Rat	PTLNHNSFASHLNHTHNVASVSDLLONATPPCQRRGGHGFRTFEDVYLFDEVH.....GCDSPOSPIPTFSGHSLDQV	1280
Alf	TRHNSLQDSGDDQAGERTDILKALIRE.....QEGISTHAAGVDSICITQ.....RDEGAPRHFSHMSSELY.....YVEPL	1257
Rat	RDEPIHLKVAITKDGIED, DRLAAMFREFQNKATLVEGIRLRLFLVAKQFRKQVCEVDQRFHREFPCFTFRARDKFEEDIT	1389
Alf	LKLEPSTLYLEKCLKCYENITRTPSRDNNML.....TVDTKQPIQ, RHLRLTLIRQPTREGSYSTQRLDAETSLQAMSTSR	1345
Rat	RKLEPALAQLEKNDKDFD, LTAIPAHNRHLYGAKYVGVETVITQIFVFRATIRHSLDYKASFEYLQWGERL.....	1448
Alf	IFRSLGARELELHSHMTIKSENAMHLYITREQIDDLVPSKINTEAGQETTYEALIEAQEHSSGVVRHRLGVFYETIK	1434
Rat	.....LLEADELVAFHNTVTOCHILFLHVPYTIQD.....PSKIEESVSHVYRIGSLRMLVQALKE	1514
Alf	.....LWITACGANGAURYVNYTHTCTVHIREMADTHKVVSSVYK, GPLNVPYHNYPLGGIDRKLAAKSTTCTYDPLA	1522
Rat	.....NRLTTTGALPI, RLPLTHESGTYDLSLKEYTDSATQINQAYGKQDGLHMLINTPYTKLQSKRQALSGTITDIPEN	1603
Alf	FTSLLE.....QSVSQDTQIRANDKLVTELKFSKAGSGVSLVPAERLPLNDVGHVAMHMETPKFSPQRTLVPSDHTFQAG	1609
Rat	FRSLKLVESHSQAFPLSPPLPSDITVTLVLDQDGG.....LVHNRHLPGEIGHVAHNSKSPPEPGDQVIVIGDITRIB	1688
Alf	SFEPEDAFFRATLACAKGIPPLTIAAGSARLGAEEVKAFCVGVASKEEPGFGVYVLPEDYARIGS.....SVAMHKLSEGT	1677
Rat	SFEPEDLFLAASLARAEGIPPLTIAAGSARLGAEEIRHGFHVAVDSQEPYKGTLYLTPODYKRYASLNSVYKWE, DGEES	1779
Alf	RNVITDYGEGDGLGVNLSGQATAGATRAYKFTFTLYVTGRTVIGATLARGHICQRLDQPIITLFGSALHKLGRVYSSHQ	1787
Rat	RYKITDYGEGDGLGVNLSGQATAGATRAYKFTFTLYVTGRTVIGATLARGHICQRLDQPIITLFGSALHKLGRVYSSHQ	1887
Alf	LGPKINATVHNYLTVSDLEGVSSILKQSVYPSVGGALPIVKKLDPPEVEYELPEN, SCDPRAATSGTLDV, NSQGLGIFDKD	1874
Rat	LGQGITVHNYHNYLTVSDLEGVSSILKQSVYPSVGGALPIVKKLDPPEVEYELPEN, SCDPRAATSGTLDV, NSQGLGIFDKD	1957
Alf	SPVETLEGAARTVYDARLGLPHILVAVETQVNIIPADPGLDSHERVYVPAQGVNPPDSATKAQALDFNREELPFIAMVRS	1984
Rat	SPSEIHPQVLEGAARTVYDARLGLPHILVAVETQVNIIPADPGLDSHERVYVPAQGVNPPDSATKAQALDFNREELPFIAMVRS	2047
Alf	FSGGQDFFETLQAGSTIVENRKYQPIFYVIMHGLRGGAVVDSRINSHDIEHTAERTAGHVEPEGHEIKFRTELLECH	2054
Rat	FSGGQDFFETLQAGSTIVENRKYQPIFYVIMHGLRGGAVVDSRINSHDIEHTAERTAGHVEPEGHEIKFRTELLECH	2137
Alf	RLDQGLTLKLESEKSNKDYGTATSDQGIQIRFRKGLLPLTQIATKFAELHDSLHKAAGVIREVLDVKSSEYVYQVLRHRIE	2144
Rat	RVDYPTIRLALRGLTPELSP, TERKEKSLKEREFTPIYVAVQADLQVTPGRMGKGVINDILDKTSRTFFVLRKLLLED	2225
Alf	SLINIVRDAAGDGLSYSAMLLKEVYNSDIAKGRDAHLDQEAFFVRDQ.....PANYDKLELRYQLLTLTIGSDALD	2226
Rat	.....LVKGIHSAPELTDGQIHLNRVFEVE, GTVYKAVVNDKLLVLEQLDEQVRSVIEENTKYSRDLVQIRSLVQAPV	2314
Alf	QALPQGLAALSKLEASRDKLISE, RYKLV.....	2257
Rat	AM, OSIVHTQHSPTQAEVYRILSTHDSPT 2345	

FIG. 2. Alignment of alfalfa and rat ACCase amino acid sequences. Biotin binding site is indicated by asterisk. Putative binding sites for ATP (amino acids 221–240), acetyl CoA (amino acids 1884–1903), and carboxybiotin (amino acids 1594–1643) are shaded.

binding site in alfalfa ACCase (Fig. 2) may be involved in the formation of a hinge region for the movement of carboxybiotin (36). Putative acetyl CoA and carboxybiotin binding sites in the alfalfa ACCase have 80% and 60% sequence identity to the corresponding regions in the rat ACCase (Figs. 2 and 3). Also, a putative ATP-binding site in the alfalfa ACCase matches the consensus sequences Gly-Xaa-Gly-Xaa-Xaa-Gly or Gly-Xaa-Xaa-Gly-Xaa-Gly shared by nucleotide binding proteins (37, 38) and has 86% identity with the corresponding region in the rat ACCase (Figs. 2 and 3). The sequence Glu-Leu-Asn-Pro-Arg located at positions 354–358 in the BC region of alfalfa ACCase (Fig. 2) is similar to the consensus sequence Glu-Met-Asn-Pro-Arg, which was proposed to be a candidate for the catalytic site of biotin-dependent carboxylases and carbamoyl phosphate synthetases (39).

**Genomic Organization of ACCase.** Two bands with different hybridization intensities were detected on probing alfalfa genomic DNA with the 3'ACC cDNA (Fig. 1) at high-stringency hybridization and washing conditions (65°C) (Fig. 4A). Southern blot analysis using mixed alfalfa 3'ACC and 5'T cDNA probes (Fig. 1) detected, at stringent hybridization and washing conditions (55°C), only one band in *Arabidopsis* (Fig. 4C), approximately two bands in rice and wheat (Fig. 4B), and three or four bands in pea, peanut, soybean, and tobacco (data not shown).

**Elicitation of ACCase.** The specific activity of ACCase in alfalfa suspension cells increased in response to treatment with yeast elicitor, reaching a maximum 4- to 5-fold increase between 8 and 12 h postelicitation (Fig. 5). The average specific activities of control and induced cells between 8 and 12 h postelicitation were 5.3 and 23.4 nmol of acetyl CoA per mg of protein per min, respectively.

ACCase transcripts were not detected in unelicited alfalfa cell suspension cultures. However, transcripts were significantly induced within 2 h of exposure to fungal or yeast elicitor (Fig. 6). ACCase transcripts reached maximum levels at 6 h and declined rapidly by 8 h after treatment with the fungal elicitor (Fig. 6A). In contrast, transcript levels remained elevated beyond 8 h in yeast elicitor-treated cultures (Fig. 6B).

## DISCUSSION

**Possible Role for Multiple ACCase Enzymes in Plants.** In most organisms, the major role of ACCase is production of malonyl CoA for fatty acid synthesis. In plants, the biochem-

### ATP

Alfalfa	G T T P A N I K A S V G G G G K O I R E V	240
<i>C. cryptica</i>	N G C I M I K A S G G G G K O I R F V	306
Yeast	G T T P A N I K A S G G G G K O I R V	265
Rat	G T T P A N I K A S G G G G K O I R V	323
<i>E. coli</i> BC	G T T P I K A S G G G G G R H R V	172

### Acetyl-CoA

Alfalfa	A R V V T T R A K L G G T P V G T V A	1903
<i>C. cryptica</i>	G K S V V I O R C R L G G I F H O A T	1777
Yeast	A K C V V O R A R L G G I P L O V I G	1897
Rat	A C V V G R A R L G G I P V G V A	1986
<i>E. coli</i> α-CT	D K A I V G G I A R L D G R F V N I C	117

### Carboxybiotin

Alfalfa	G R T I L V S N D V T F K A S F P R E D A F	1618
<i>C. cryptica</i>	G R Q V V I V D V T V S S H F G V E D E V	1500
Yeast	G R Q V V V A M D I T F N I S F G P Q E D E F	1603
Rat	G R D V I V I C H D I Y R I L O S T F Q E D L	1697
<i>E. coli</i> β-CT	G H P V A A A F E F A H G S H G S V G C A R	140

Alfalfa	F R A V T D L A C A K E I F L I T L A H S G A R	1643
<i>C. cryptica</i>	F K A S K Y A R E N L F R V Y I A C H S O A R	1525
Yeast	F N K V T E A R K R G I F R T L A H S O A R	1628
Rat	F L R A S E L A R E D E I F R I T V A H S O A R	1722
<i>E. coli</i> β-CT	F V R A E Q A L E D N C F L I C F S A S C A A	165

FIG. 3. Amino acid sequence alignments of putative ATP, acetyl CoA, and carboxybiotin binding sites from several ACCases. Residues identical to the putative alfalfa sequence are shaded. α-CT, α subunit of CT; β-CT, β subunit of CT.

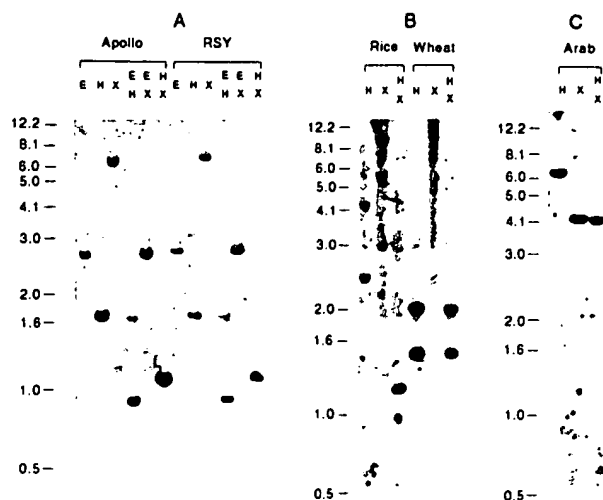


FIG. 4. Genomic organization of ACCase in alfalfa cv. Apollo and Regen SY (RSY), rice, wheat, and *Arabidopsis* (Arab). Genomic DNA (10  $\mu$ g) was digested with *Eco*RV (E), *Hind*III (H), and *Xba*I (X), or mixtures of these enzymes, resolved on a 0.7% agarose gel, and blotted to GeneScreenPlus. Blot A was probed with the 3' ACC cDNA. Blots B and C were probed with a mixture of the 3' ACC and 5' T cDNAs. Numbers on left are kb.

ical fate of malonyl CoA is considerably more diverse. ACCase activity has been reported in plastid preparations from numerous species and it is likely that malonyl CoA is used almost exclusively for *de novo* fatty acid synthesis in this organelle (3). Outside the plastid, malonyl CoA enters into several pathways (7–9, 11, 40), including the elongation of fatty acids (6). Production of malonyl CoA for these diverse roles could be accomplished by multiple ACCase isozymes in different subcellular compartments or, alternatively, by a single enzyme accompanied by transport of malonyl CoA or malonate across membranes.

Early observations on fatty acid elongation in oilseeds strongly implicated the participation of multiple ACCase

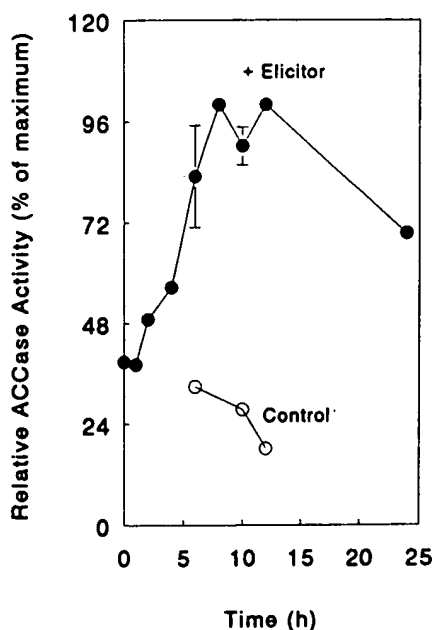


FIG. 5. Elicitation of ACCase activity in alfalfa cell suspension cultures in response to yeast elicitor (50- $\mu$ g glucose equivalents per ml of medium) (+ elicitor) or an equivalent volume of water (control). Bars represent the spread of values from duplicate elicitor-treated cultures.

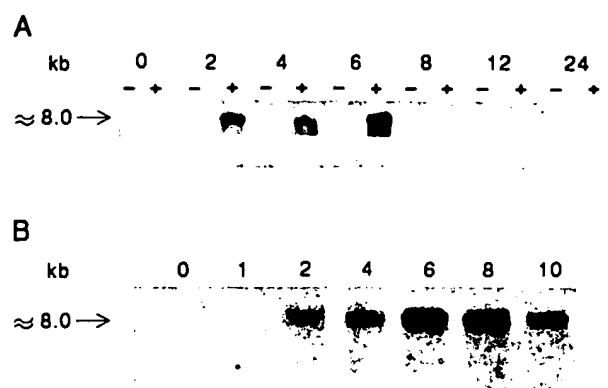


FIG. 6. Expression of ACCase transcripts in total RNA (15  $\mu$ g) isolated from alfalfa cell suspension cultures at the indicated times (hours) after exposure to *Colletotrichum* elicitor (50- $\mu$ g glucose equivalents per ml of medium) (lanes +) or an equivalent volume of water (lanes -) (A) or yeast elicitor (50- $\mu$ g glucose equivalents per ml of medium) (B). Blot A was probed with the 3' ACC cDNA. Blot B was probed with a mixture of 3' ACC and 5' T cDNAs.

enzymes in the synthesis of very long chain fatty acids (41). Recently, further evidence for more than one form of ACCase has been obtained from maize leaves, where both herbicide-tolerant and -sensitive isozymes could be resolved (5, 42), and from pea chloroplasts (23), where a transcarboxylase subunit of ACCase has been characterized. Thus, although the current evidence is indirect, the most likely interpretation of these data is that plants contain ACCase isozymes in both the plastid and the cytosol.

Two major lines of evidence suggest that the cDNA described in this work represents a cytoplasmic form of ACCase. First, the N-terminal amino acid sequence does not appear to code for a chloroplast transit peptide or for any other signal sequence. All nuclear-encoded plastid proteins contain an N-terminal extension of  $\approx 50$  amino acids, which directs transport of these proteins into the plastid. When aligned, rather than containing an N-terminal extension, the N-terminal sequence of the alfalfa BC domain is  $\approx 20$  and  $\approx 80$  amino acids shorter than the yeast and chicken BC domains, respectively. In addition, a 16-residue sequence, which is highly conserved with all other eukaryotic cytoplasmic ACCases, begins 36 residues into the alfalfa sequence. Second, ACCase activity and transcripts are strongly induced by treatment of the alfalfa cell cultures with yeast or fungal elicitors. This induction is characteristic of the chalcone synthase enzyme, which requires malonyl CoA as a cosubstrate for biosynthesis of flavonoid in the cytoplasm (7–9). Direct demonstration of the subcellular location of the protein associated with this ACCase cDNA will await immunocytochemical analysis or chloroplast uptake studies of *in vitro* translated ACCase cDNA.

At this time it is not clear how closely related are the cytoplasmic and plastid forms of ACCase. Partial cDNA sequences have been reported for maize and wheat ACCase in the GenBank data base but it is not certain whether these code for plastid isozymes. These monocot sequences have  $\approx 65\%$  identity (80% similarity) with the dicot alfalfa ACCase in their overlapping regions. In contrast, we have sequenced an *Arabidopsis* genomic clone for ACCase that is  $>80\%$  identical (89% similar) to alfalfa in amino acid sequence (43). More detailed sequence comparisons indicate that all the available plant ACCase sequences have local regions of very high identity and therefore heterologous probes at low stringency may cross-hybridize between them. For this reason, hybridization strategies to determine tissue or developmental expression patterns may need to account for potential contributions from a mixture of ACCase mRNA species.



**Genomic Organization.** The detection of two bands in Southern blots with different hybridization intensities suggests that ACCase in alfalfa is encoded by at least two related genes. Southern blot analysis of alfalfa ACCase-related genes in other plants suggests that *Arabidopsis*, rice, and wheat may have one or two corresponding genes in their genome. At this time it is not possible to determine whether genes coding for both cytoplasmic and plastidial forms of ACCase would be detected by the alfalfa probe in these blots. Since the nucleotide sequences of >71 independently isolated alfalfa ACCase cDNAs were identical in their overlapping regions, both in the coding and in the untranslated regions, it is likely that all characterized cDNAs isolated from the elicitor-induced alfalfa library correspond to one gene.

**Elicitor Induction of ACCase Transcripts and Activity.** The elicitor-induced increase in ACCase transcript levels (Fig. 6) precedes the increase in enzymatic activity (Fig. 5), and this timing correlates with the induction kinetics of transcripts/activities of the series of enzymes involved in the synthesis of the isoflavonoid phytoalexin medicarpin from phenylalanine (9). It is possible that most of the undetectable ACCase message in unelicited, control alfalfa cell cultures is associated with the plastidial form, which may not cross-hybridize with the cytosolic ACCase probe (Fig. 6).

Elicitation of isoflavonoid phytoalexin biosynthesis in legume cell suspension cultures requires three molecules of malonyl CoA as cosubstrate with 4-coumaroyl CoA in the chalcone synthase reaction (9). Thus, ACCase is coinduced with chalcone synthase in elicited alfalfa and soybean (8) cell suspensions, suggesting that insufficient malonyl CoA is produced under normal conditions to satisfy the requirements for both flavonoid synthesis and fatty acid elongation. ACCase is also induced by UV irradiation in parsley cells, associated with the accumulation of UV-protective flavonoid compounds (7); in this species, fungal elicitation does not involve ACCase induction (42) as the parsley furanocoumarin phytoalexins do not require malonyl CoA for their synthesis. It is not yet known whether elicitation and the concomitant induction of ACCase has quantitative or qualitative effects on fatty acid elongation. The present work opens up the possibility of investigating regulation of channeling of malonyl CoA into primary and secondary metabolism by genetic manipulation of ACCase levels.

We thank Karen Dalkin for construction of the alfalfa cDNA library, Nancy Paiva for providing some of the elicited cell suspension cultures, Gansan Gowri for assistance in constructing the alfalfa genomic library, and Yves Poirier and Dave Shintani for critical reading of the manuscript.

- Post-Beittenmiller, D., Jaworski, J. G. & Ohlrogge, J. B. (1991) *J. Biol. Chem.* **266**, 1858–1865.
- Post-Beittenmiller, D., Roughan, G. & Ohlrogge, J. B. (1992) *Plant Physiol.* **100**, 923–930.
- Harwood, J. L. (1988) *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **39**, 101–138.
- Kim, K.-H., Lopez-Casillas, F. & Bai, D.-H. (1989) *FASEB J.* **3**, 2250–2256.
- Egli, M. A., Gengenbach, B. G., Gonwald, J. W., Somers, D. A. & Wyse, D. L. (1993) *Plant Physiol.* **101**, 499–506.
- Pollard, M. & Stumpf, P. (1980) *Plant Physiol.* **66**, 649–655.
- Ebel, J. & Hahlbrock, K. (1977) *Eur. J. Biochem.* **75**, 201–209.
- Ebel, J., Schmidt, W. E. & Loyal, R. (1984) *Arch. Biochem. Biophys.* **232**, 240–248.
- Dixon, R. A. & Harrison, M. J. (1990) *Adv. Genet.* **28**, 165–234.
- Lanz, T., Tropf, S., Marner, F.-J., Schroder, J. & Shroder, G. (1991) *J. Biol. Chem.* **266**, 9971–9976.
- Kionka, C. & Amrhein, N. (1984) *Planta* **162**, 226–235.
- Li, S.-J. & Cronan, Jr., J. E. (1992) *J. Biol. Chem.* **267**, 855–863.
- Li, S.-J. & Cronan, Jr., J. E. (1992) *J. Biol. Chem.* **267**, 16841–16847.
- Kondo, H., Shiratsuchi, K., Yoshimoto, T., Masuda, T., Kitazono, A., Tsuru, D., Anai, M., Sekiguchi, M. & Tanabe, T. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 9730–9733.
- Alix, J.-H. (1989) *DNA* **8**, 779–789.
- Takai, T., Yokoyama, C., Wada, K. & Tanabe, T. (1988) *J. Biol. Chem.* **263**, 2651–2657.
- Lopez-Casillas, F., Bai, D.-H., Luo, X., Kong, I.-S., Hermodson, M. A. & Kim, K.-H. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 5784–5788.
- Al-Feel, W., Chirala, S. S. & Wakil, S. J. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 4534–4538.
- Roessler, P. G. & Ohlrogge, J. B. (1993) *J. Biol. Chem.* **268**, 19254–19259.
- Gornicki, P., Scappino, L. A. & Haselkorn, R. (1993) *J. Bacteriol.* **175**, 5268–5272.
- Nikolau, B. J. & Hawke, J. C. (1984) *Arch. Biochem. Biophys.* **228**, 86–96.
- Kannangara, C. G. & Stumpf, P. K. (1972) *Arch. Biochem. Biophys.* **152**, 83–91.
- Sasaki, Y., Hakamada, K., Suama, Y., Nagano, Y., Furusawa, I. & Matsuno, R. (1993) *J. Biol. Chem.* **268**, 25118–25123.
- Dalkin, K., Edwards, R., Edington, B. & Dixon, R. A. (1990) *Plant Physiol.* **92**, 440–446.
- Schumacher, H. M., Gundlach, H., Fiedlar, F. & Zenk, M. H. (1987) *Plant Cell Rep.* **6**, 410–413.
- Bradford, M. M. (1976) *Anal. Biochem.* **72**, 248–257.
- Shorrosh, B. S. & Dixon, R. A. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 10941–10945.
- Millar, S. J., Dempsey, D. L. & Dickinson, P. D. (1992) *Bio-Techniques* **13**, 554–562.
- Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **83**, 8073–8077.
- Sanchez-Pescador, R. & Urdea, M. S. (1984) *DNA* **3**, 339–343.
- Chomczynski, P. & Sacchi, N. (1987) *Anal. Biochem.* **162**, 156–159.
- Shorrosh, B. S. & Dixon, R. A. (1992) *Plant J.* **2**, 51–58.
- Shorrosh, B. S. & Dixon, R. A. (1992) *Plant Mol. Biol.* **18**, 1189–1190.
- Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY).
- Kozak, M. (1987) *Nucleic Acids Res.* **15**, 8125–8132.
- Samols, D., Thornton, C. G., Murtif, V. L., Kumar, G. K., Haase, F. C. & Wood, H. T. (1988) *J. Biol. Chem.* **263**, 6461–6464.
- Sternberg, M. J. E. & Taylor, W. R. (1984) *FEBS Lett.* **175**, 387–392.
- Moller, W. & Amons, R. (1985) *FEBS Lett.* **186**, 1–7.
- Konodo, H., Shiratsuchi, K., Yoshimoto, T., Masuda, T., Kitazono, A., Tsuru, D., Anai, M., Sekiguchi, M. & Tanabe, T. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 9730–9733.
- Liu, Y., Hoffman, N. E. & Yang, S. F. (1983) *Planta* **158**, 437–441.
- Ohlrogge, J. B., Pollard, M. R. & Stumpf, P. K. (1978) *Lipid* **13**, 203–210.
- Howard, J. L. & Ridley, S. M. (1990) *FEBS Lett.* **261**, 261–264.
- Roesler, K. R., Shorrosh, B. S. & Ohlrogge, J. B. (1994) *Plant Physiol.*, in press.

## The gene and the RNA for the precursor to the plastid-located glycerol-3-phosphate acyltransferase of *Arabidopsis thaliana*

Ikuo Nishida, Yasushi Tasaka, Hideaki Shiraishi and Norio Murata  
National Institute for Basic Biology, Myodaiji, Okazaki 444, Japan

Received 26 May 1992; accepted in revised form 18 September 1992

**Key words:** *Arabidopsis*, chilling sensitivity, glycerol-3-phosphate acyltransferase, glycerolipid biosynthesis, RNA

### Abstract

The gene and the RNA from *Arabidopsis thaliana* for the plastid-located glycerol-3-phosphate acyltransferase (GPAT; EC 2.3.1.15) and their encoded product have been studied. The gene (designated *ATS1*) was isolated by screening a  $\lambda$ DASH genomic library for cross-hybridization with a radiolabeled probe prepared from cDNA for GPAT from squash. cDNA clones representing the mRNA were isolated by screening a  $\lambda$ ZAPII cDNA library for hybridization with a radiolabeled probe prepared from a DNA fragment of *ATS1*. The nucleotide sequences of the gene and the cDNA were determined, and the 5' end of the RNA was mapped by primer extension. Sequences similar to the TATA box, polyadenylation sequences and intron-splicing sequences were found at the expected locations. The pre-mRNA was 3288 nucleotides long and contained 5' and 3'-untranslated sequences of 57 and 442 nucleotides, respectively. The coding sequence of 1377 nucleotides was interrupted by 11 introns of 1412 nucleotides in total and the 3'-untranslated sequence contained another intron of 94 nucleotides. The open-reading frame encoded a polypeptide of 459 amino acid residues, the amino acid sequence of which was highly homologous to those of precursors to plastid-located GPATs from squash and pea. The enzymatic activity of a gene product that was over-produced in *Escherichia coli* confirmed the identity of the gene.

**Abbreviations:** ACP, acyl carrier protein; GPAT, glycerol-3-phosphate acyltransferase; IPTG, isopropyl- $\beta$ -thiogalactopyranoside.

### Introduction

In all organisms higher than the eubacteria, glycerol-3-phosphate acyltransferase (GPAT; EC

2.3.1.15) catalyzes the first reaction in the synthesis of glycerolipids *de novo*, namely, synthesis of 1-acylglycerol 3-phosphate (lysophosphatidic acid) from glycerol 3-phosphate and acylth-

The nucleotide sequences reported herein for the gene and the mRNA will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession numbers D00672 and D00673, respectively. The designation of the gene, *ATS1*, has been changed from *atsA*, which was used in our preliminary report [22] of this work, in accordance with the recommended nomenclature for genes in *Arabidopsis*.



ioesters. Therefore, this enzyme is indispensable for the biosynthesis of membrane lipids [18]. In cells of higher plants at least three types of GPAT, which differ in their subcellular location, contribute to biosynthesis of glycerolipids. They also differ from one another in terms of their substrate specificity and binding to membranes: GPAT bound to membranes of the endoplasmic reticulum is specific for acyl-CoA [3]; plastid-located GPAT is soluble in the stroma and uses acyl-(acyl-carrier protein) (acyl-ACP) as substrate [4]; and mitochondrial GPATs appear to vary in their binding to membranes among plant species and they react with both acyl-CoA and acyl-ACP [5]. Molecular cloning of genes for plant GPATs is important if we are to understand the regulation of biosynthesis of glycerolipids in plant cells.

The plastid is the sole site of biosynthesis of fatty acids in plant cells [31]. The final product of the synthesis in the plastids is acyl-ACP [32]. The acyl group of acyl-ACP is either directly incorporated into glycerolipids within plastids or, after hydrolysis, it is exported from plastids for biosynthesis of glycerolipids in the endoplasmic reticulum and mitochondria [28]. Plastid-located GPAT competes for acyl-ACP with another stromal enzyme, acyl-ACP hydrolase [23], which hydrolyzes the acyl-ACP to fatty acids and ACP. Thus, the relative activities of the transferase and the hydrolase have been postulated to play a regulatory role in determining the fate of fatty acids in the biosynthesis of glycerolipids in plant cells [15, 23]. Another interesting aspect of the plastid-located GPAT is its unique specificity toward substrates. We have demonstrated that the fatty-acid selectivity of the GPAT is a major determinant both of fatty-acid unsaturation of phosphatidylglycerol from plastidial membranes and of the chilling sensitivity of plants [19, 20].

We have found three isomeric forms of plastid-located GPAT in squash cotyledons and have purified two of the isoforms to homogeneity [21]. A cDNA clone for one of the purified isoforms has been obtained by screening a squash cDNA library for immuno-cross-reactivity. The nucleotide sequence of the cDNA has been determined [9].

*Arabidopsis thaliana* provides a simple system for molecular cloning of plant genes [25]. The genetic analysis of mutants defective in plastid-located GPAT activity suggests that the GPAT gene is a single-copy gene in the *Arabidopsis* genome [13]. This plant system seems to be very well suited for studies of the sequences of the gene, the RNA and the protein responsible for plastid-located GPAT activity. We report here the cloning of the gene and the cDNA for a precursor to GPAT from *A. thaliana* (the gene is designated *ATN1*), and we discuss features of the gene, the RNA and the encoded amino acid sequence. The identity of the gene is confirmed by the extent of sequence homology with GPAT from squash [9] and pea [33] and by the enzymatic activity of a gene product that is over-produced in *E. coli*.

## Materials and methods

### *Plant materials and bacterial strains*

The Landsberg *erecta* strain of *A. thaliana* was obtained from Dr A.R. Kranz (Botanisches Institut, J.W. Goethe-Universität, Frankfurt am Main, FRG). *Arabidopsis* seeds were germinated on peat plates irrigated with 1/1000-strength Hyponex solution (Hyponex, Miami, FL, USA), and the plates were incubated for 3–4 weeks at 25 °C under continuous fluorescent illumination (20  $\mu\text{E m}^{-2} \text{s}^{-1}$ ). Leaves were used in further experiments. *E. coli* BL21 (DE3) cells [30] and the translation plasmid pET-3c [27] were generous gifts from Dr F.W. Studier (Biology Department, Brookhaven National Laboratory, NY, USA) and were obtained through Professor Y. Shimura (Department of Biophysics, Kyoto University).

### *General methods for manipulation of nucleic acids*

Extraction of nucleic acids from *A. thaliana*, DNA/DNA and RNA/DNA hybridization analysis, and other techniques for manipulation of nucleic acids were performed by the procedures

described by Ausubel *et al.* [1]. Poly(A)<sup>+</sup> RNA was purified by adsorption to Oligotex dT30 resin (Nippon Roche, Tokyo). Radiolabeled probes were prepared by nick translation [26] using [ $\alpha$ -<sup>32</sup>P]dCTP (111 TBq/mmol; DuPont/NEN Research Products, Wilmington, DE, USA). Excision and recircularization of pBluescript (sk<sup>-</sup>) plasmids that contained cloned cDNA inserts were carried out *in situ* using the helper phage R408 (Stratagene, La Jolla, CA, USA) in accordance with the manufacturer's protocol.

### Construction of recombinant libraries

Construction of *Arabidopsis* genomic libraries followed the procedures described by Ausubel *et al.* [1]. Partially *Sau*3AI-digested fragments of *Arabidopsis* DNA were fractionated by centrifugation on a sucrose gradient and fragments of about 20 kb in length were ligated with *Bam* HI-digested arms of a phage vector,  $\lambda$ DASH (Stratagene). An *Arabidopsis* cDNA library was constructed by the manufacturer's protocol which was supplied with the cDNA synthesis kit (Pharmacia P-L Biochemicals, Uppsala, Sweden). cDNA was synthesized by priming with oligo(dT)<sub>12-18</sub> primers and ligated with *Eco* RI-digested arms of another phage vector,  $\lambda$ ZAPII (Stratagene) through *Eco* RI/*Not* I linkers. DNA concatemers were packaged into phage particles with the packaging extract, Gigapack II Gold (Stratagene). The genomic libraries were amplified but the cDNA library was not amplified.

### Screening of recombinant libraries

Phages were plated on agarose plates (10 cm  $\times$  14 cm) at a density of less than 20 000 plaque-forming units per plate. Plaques were lifted onto nylon membranes (GeneScreen Plus; DuPont, Boston, MA, USA), and the membranes were then rinsed in 5  $\times$  SSC (1  $\times$  SSC is 0.15 M NaCl, 0.015 M sodium citrate pH 7.0, 0.5% SDS, 1 mM EDTA) at 42 °C for 1 h. Hybridization was preceded by prehybridization in the ab-

sence of radiolabeled probes at 65 °C for 1 h. Hybridization cocktails contained 6  $\times$  SSC, 5  $\times$  Denhardt's solution (1  $\times$  Denhardt's solution is 0.02% Ficoll, 0.02% polyvinylpyrrolidone, 0.02% bovine serum albumin), 10% dextran sulfate, 0.1% SDS and 100  $\mu$ g/ml salmon testis DNA. A squash cDNA probe was prepared from the insert DNA of a cDNA clone for squash GPAT, AT03 [9]; an *Arabidopsis* DNA probe was prepared from the 2.6 kb *Bam* HI restriction fragment of an isolated genomic clone,  $\lambda$ 6; and an *Arabidopsis* cDNA probe was prepared from the 1.3 kb *Sal* I-*Sty* I restriction fragment of an isolated cDNA clone, AR1 (Fig. 1B). Cross-hybridization with the squash cDNA probe was carried out at a probe concentration of 10<sup>5</sup> cpm/ml and at 50 °C with washes in a washing solution that contained 2  $\times$  SSC and 0.1% SDS at 40 °C, whereas hybridization with either the *Arabidopsis* DNA probe or the *Arabidopsis* cDNA probe was carried out at a probe concentration of 5  $\times$  10<sup>5</sup> cpm/ml, at 65 °C, with washes in the same washing solution at 65 °C.

### Determination of DNA sequences

DNA sequences were determined by the dideoxy chain-termination method [29] with deoxy-7-deazaguanosine triphosphate used in place of dGTP [17]. Oligonucleotide primers were synthesized with a DNA synthesizer (model 391; Applied Biosystems Japan, Osaka).

### Over-production of a GPAT fusion protein in *E. coli*

A 1.2 kb *Hga* I-*Eco* RI restriction fragment of an isolated cDNA clone for GPAT from *Arabidopsis* (AR1 in Fig. 1B) was blunt-ended and subcloned into the *Bam* HI site of pET-3c to obtain the plasmid pET-AR1. In this way, a new open reading frame, encoding a fusion protein between a T7 bacteriophage leader sequence of 12 amino acid residues, which included the junction sequence, and the sequence encoding GPAT from *Arabidopsis* of 369 amino acid residues, was con-

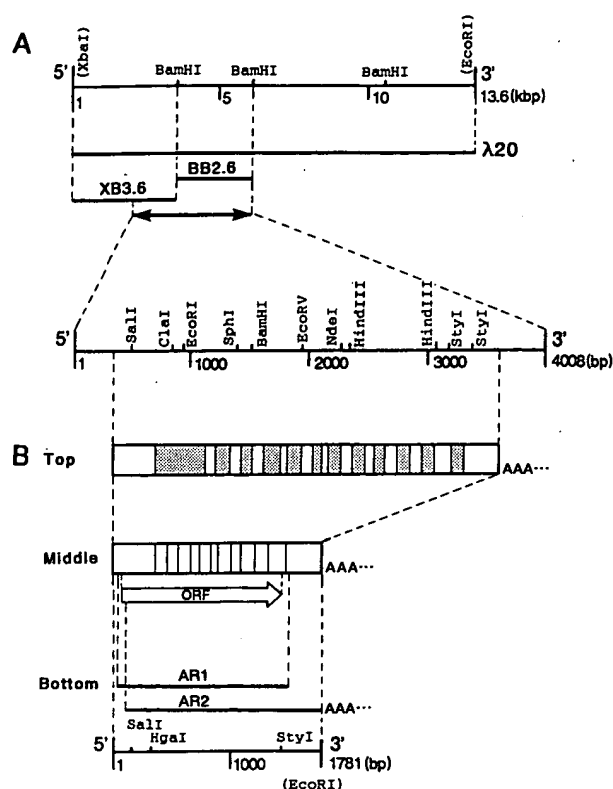


Fig. 1. Structures of the gene (*AT51*) and the mRNA for plastid-located glycerol-3-phosphate acyltransferase from *A. thaliana*. A. Restriction maps of isolated genomic clones. The nucleotide sequence of a DNA segment, indicated by a bar with arrows at both ends, was determined with nested deletions [1] of XB3.6 and BB2.6. Restriction enzymes in parentheses indicate that the restriction sites are located within the linker sequences of the vector arms. B. Structures predicted for the pre-mRNA (top) and the mRNA (middle) are compared with those determined for isolated cDNA inserts (bottom). Exons and introns are shown by open and closed boxes, respectively. The open reading frame is designated as ORF and the direction of translation is indicated by the arrowhead. Nucleotide sequences of cDNA inserts, AR1 and AR2, were determined by use of synthetic oligonucleotide primers. The 5' end of the mRNA was estimated by primer extension (Fig. 3) and the number 1 was assigned to the first nucleotide of the mapped 5' end. The cDNA inserts AR1 and AR2 start from nucleotides 43 and 64, respectively.

structed under the control of a T7 bacteriophage promoter. Over-production of the GPAT fusion protein in *E. coli* BL21 (DE3), after transformation with pET-AR1, was achieved by following the procedures described by Hoey *et al.* [8]. Cells were suspended in 20 ml of a solution that contained 45 mM Tris-HCl pH 7.4, 2 mM DTT, 10 mM sodium ascorbate, 10% (v/v) glycerol, 1 mM benzamidine hydrochloride, 5 mM 6-aminohexanoic acid and 0.024 mM leupeptin, and were broken by passage through a chilled French pressure cell (FA-030; SLM Instruments, Urbana, IL, USA) operated at 57 MPa. The homogenate was clarified by centrifugation at  $16\,000 \times g$  for 10 min and then at  $100\,000 \times g$  for 60 min to obtain the water-soluble protein extract.

#### Assay of GPAT activity and analysis of proteins

GPAT activity was measured as described previously [21]. SDS-PAGE was carried out as described by Laemmli [14].

## Results

#### Isolation of a gene from *Arabidopsis* for a precursor to plastid-located GPAT

An *Arabidopsis* genomic library was screened for cross-hybridization with the squash cDNA probe. From 18000 recombinant clones, two positive clones were obtained and designated λ6 and λ10. These positive clones contained a 2.6 kb *Bam* HI restriction fragment of *Arabidopsis* DNA which hybridized with the squash cDNA probe. However, when partial DNA sequences were determined, they revealed that the positive clones were not full-length clones. To obtain segments of DNA that included the entire length of the gene,

Fig. 2. Nucleotide sequences of the *AT51* gene and the 5' and 3' flanking sequences. The deduced amino acid sequence is shown in the one-letter code and numbered separately. Exon sequences are shown in brackets. A TATA-like sequence and a putative polyadenylation signal are indicated by single and double underlinings, respectively.

1 TATTCAATGTTGTACACAACAACATGACTACTCTCAAACAACAACAAGGCCAAATACAATGTACGGCAGCACTAACAAAAATAAATA  
91 AAAGATGTGAGGTTATAATTTGGTCTTGAGTCTTCTTACTCACAAAGTCACAATATCTAAATTTTATTAAATTTAAAATTAATACATAG  
181 TATCATTTATAATTTAGAGTTCTTCAATGAATAAGTTTACTTAGGGGAAAATTAATTAATTTAAAAAAGAGTATTTAAATTTGGTG  
271 AAAAATAAAAAAGAAGAACCGTGAAAGAGGGGACTACGGGTAATAAAGGAAGGAAGAGATCTCCGCTCACACACATTAACACTC  
361 CACACGCTCTGCTTGCACCAATCACCAACACGCTTAAATGACTCTCACGTTTTCTCTCCGCGCAACCGTTGCCGTTGCTGCTGCAA  
1' M T L T F S S A A T V A V A A A  
451 CCCTAACCTCTCCGCTAGGGTTCCGGTTTATCCAAGTCTCGTCTCGTCTGAGTATAGTATCTTTGAGTTAAACCGGAGAAGC  
18' T V T S S A R V P V Y P L A S S T L R G L V S F R L T A K K  
541 TGTTCTGCGCGCTCTTCGTTCTCGCGCGGCGTTAGTGTGAGCGCATGCTGAGCTAGTTCAAGATAAAGAAATCGTCCGCTCGCGCGG  
48' L F L P P L R S R G G V S V R A M S E L V Q D K E S S V A A  
631 GCATTCGCTTCAATGAAGCGCGGTGAGACGCGGAGTGAGCTTAATCATTCCCCTACTTTCTTGGATGCGCGAAGTGAACAAGGTTAGT  
78' S I A F N E A A G E T P S E L N H S R T F L D A R S E Q  
721 TACTCTTCTCGTCTTCTTCTGAGGAATTTGAATTGACTGTGTGTAGCAATTCGGAATGGAATTTAGAGTTGTGGAGATTTGTGCA  
811 TCCTTTTGTCTCATAATCTTATCGCGAAATCGATATTTGAGTTGTTCTGTTGTGACTGGATAAGGTCATTGAATGACTTCTTGCT  
901 TTGCTCTGATCTTGCTCCATTCTTGTTAAATGGATAAGACGCTTTTGGAAATCTTAGGTTGTGTGCTCTTGATATGCTTTGCTGTT  
991 CTTTAAGAAAGATCAGTTTCATAAGGATGCAATCAGATAATTTCTCTTTTCAAAAAGTCAAGTAGGAGTTCTGAGTGATGTGTAACAA  
1081 TGCTTGATTAACATTTGTTTATCTTCATTTCATCGTCACTCTTTTATCTGTTATCAAGAAGAGGCTGAGGAGGAGGTTGCCAG  
106' D L L S G I K K E A E A G R L P  
1171 CAAATTTGCGACGAATGAAGAATTTGATTTGGAACATCAAAAATGCAAGTAAATGAATCCTTTTATCACTCACTAGTTTGTACGATTT  
122' A N V A A G M E E L Y W N Y K N A  
1261 AGTGCTATGCTTTGGAATTAATTAGATTAAAGAACATCAGTCCACTGAACTTTGGCATTGTTGTCTTACTGATGGTTTTAGGTTTAAAG  
139' V L S  
1351 TAGTGGAGCTTCCAGGGCAGATGAACTGTTGTATCAACATGTCTGTGCTTTTGTATCGATGCTTCTTGGTGGAGTACTTCTTCT  
142' S G A S R A D E T V V S N M S V A F D R M L L G V E  
1441 GTCAATCTCTTGAATGATTTTTCTACTAGGAGATGTCATCTAAAAATCTTTTACTTGGAGCTTCTTGGCTGTGAATATTTCAGTAT  
168' D  
1531 CTTATACTTTAATCCATATATAAAGCAGTCAGAGAACCATTGACTACATACATTTTGGCCATCATCATCCGCTCTTATTGATTT  
169' P Y T F N P Y H K A V R E P F D Y Y M F V H T Y I R P L I D  
1621 TCAAGTTAAGCTGGAATATGCACTCATAGTTATCAGATTTTATTCTAACTGAATGTAAATAGAAAGCAATATCGTTTGGTTAGTATCATG  
199' F K  
1711 GATCTCCAGTTTTCTTCTGCCATCAAGTGTTCATCTACTAACAAACCTTGAATATGTTATGCAGTAAATCGTAGCTTGGAAATGCTTC  
201' N S Y V G N A S  
1801 TATATCTCTGAGCTGGAAGACAAGATTCGACAGTTCATCTTCCCTCTTCTTGTGCTCTAGTTAATGCTGTTTCTTTCATCAACAGTT  
209' I F S E L E D K I R Q  
1891 AAAGTCTCTTCCATTGAGATTTTAATACTCAACTGTTTAAATACAGGACCAATATCGTGTGATATCAAAACATCAAGTGAAGCTG  
220' G H N I V L I S H Q S E A  
1981 ATCCGGCTGTATTTCTCTATTGTCTGAAGCACAATCTCCTTCTAGGAGAGAACATTTGAGCCCTCGAGCCTTGCTTCTGAGCTATTA  
234' D P A V I S L L L E A Q S P F I G E N I  
2071 TAGCTATTGTCTCTGTTCTTATTGTTCTTGTGGTGCAGAAATGTGTGGCTGGTGTGAGTCATCACTGATCCCTCTTGTAAAGCGGT  
254' K C V A G D R V I T D P L C K P  
2161 TCAGTATGGGAAGTATCACGAGCTTTCATTGTCATGGTATGCTACCTGATGGGCTGAAGTGAATCTTAGCACCTTGATGACTTCCG  
270' F S M G R  
2251 GGATTTCTACTGCTTTCAGTAACTCATATGTTTACTCGAAAAGCACATGAATGTTGATCCTGAGCTTGTGACATGAAAAGAAAAG  
275' N L I C V Y S K K H M N V D P E L V D M K R K  
2341 CAAACACACGAAGCTTAAAGGAGATGGCTACAATGCTAAGTAAATGGAATCAATAGCGAGTGTCTTATTGATCTCATAGGAACAGA  
298' A N T R S L K E M A T M L R  
2431 AATTAAGGAAAACATGATGATGATTTAGTCTGCTAATGTTGATGAACCCCTGTAGTCTGCGGCGTCAACTTATATGAGTTGCACCAAGCGGTG  
312' S G G Q L I W I A P S G  
2521 GAAGGGACCGCCGAATCTTCTACTGGGGAATGTTCTCTTAAGTATTTGATGGAATACAGAGATCTTATCTGAGTCTGGTAGATAT  
324' G R D R P N P S T G G E W F P  
2611 GCAAGTACTAATCTAGTAATGTTGAGTCAACCTTTGATGCTTCTCGGTAGACAACATGAGAAGACTGGTTGAACATTTGCGGCTCTCGT  
338' A P F D A S S V D N M R R L V E H S G A P  
2701 GACATATATCCAATGCTTTTGCTTTGATGACATGCCCTCCACCCGATTTGATTTGTTTCTTCTCTGTCTCTCTCTCT  
359' G H I Y P M S L L C Y D I M P P P P Q  
2791 CCGTTTGGGCTAACTCTGGTGTACTTCTCTCATCATCACTTCAITGTGCATCTGCGAGTCTTGAGAAAAGAAATCGGAGAGAAAAGATT  
378' V E K E I G E K R L  
2881 AGTTGGGTTTCAAGTACTGGAATATCAATGTCTCTGAAATCAACTTCTCAGACGTCACAGCAGACTGCGAGAGCCCTAATGAGTCTGT  
388' V G F H G T G L S I A P E I N F S D V T A D C E S P N E  
2971 TCTGTAACACTCAGCACCAATGACTTTAGTTTCAAGATCGAGTAATTCATCATGCATATAAATACATTCTTGGTATGCGCAGGAA  
416' A K  
3061 AGAAGCATACAGCCAGCTTTGTACAAGTCGGTGAATGAACAATACGAGATCTTAAACTCTCGGATTAACACAGAAGGAGTAGAAGC  
418' E A Y S Q A L Y K S V N E Q Y E I L N S A I K H R G V E A  
3151 ATCAACTCTAAGGTCTCTTGTGCAACCTTGAATAGTCTCTGTTTTAGGCTTAAACACTTCAAAACTCATAAATCTTCTGCTCAG  
448' S T S R V S L S Q P W N \* 459'  
3241 AAGTTTGTGTGCAATGATATATATATGAGAGAGAGAGAGATGTTCTTTTCTTATTTGCAAGTACACAAACCAATCGGAAAAATCACTCA  
3331 AAAAAAAAAAGGAGATATATTTACACGTTGCTGAGGATGTTTTTTTTTTTTTTTCAACAGAGTCCAAAGCTCCTTGAACAACTCAAG  
3421 AAAAAAAAAAGAGGTACGTTGTGTTGTTTCTAAGTGAAGAGAGAGAGTTGTTTGTGACATATGAATAGTGAGAGAAAAAAGTCT  
3511 CTGCTAGGCTCTTAGATCTCATATAGTTTCTATATGAATATCTTTTCTTTTGGAGTTTCTTACTATATAGAGCCGATAATAT  
3601 ACAATCCCAATTTATGCCTCTCTCTGCTGCTTCTCATATGCTTCTATGATTTTCAATTTAGTTTTCATATGATGATTTTATGAT  
3691 GCTTGATATATCAGTCAAGACAATTTTGGGAAAATGGTTTTGTTCAATTTATACACCTTATTAGTGACAATGACAAACAGTAT  
3781 CTCTTATAATTTTGCAGAAAAAATCTGAGAAACAATTTCAACATTGATCAAAAGTGTTCAGAAATAAAGAGATTATTGGTCAAAACAT  
3871 TCCATTTTCACTATTGAAACAATTTTGTGTGATCTTAAATTTGAAGAGCGGTGAGAGATGTTTTGTGTGTTGGGTACTATTTTGTGT  
3961 TACTTGGCAAGTTGACAGAGATGATGATGCTTGGGCTTTTGGATATCC 4008

another library was screened for hybridization with an *Arabidopsis* DNA probe that had been prepared from the 2.6 kbp *Bam* HI restriction fragment of  $\lambda$ 6. After screening of 18 000 recombinant clones, five positive clones were obtained. Their restriction digests revealed that they were overlapping, and all contained the 2.6 kb *Bam* HI restriction fragment. One of the clones, which contained a 13.6 kb DNA insert ( $\lambda$ 20 in Fig. 1A), was analyzed further.  $\lambda$ 20 was digested with *Bam* HI and *Xba* I, and 2.6 kb and 3.6 kb restriction fragments were obtained that were designated BB2.6 and XB3.6, respectively. The nucleotide sequences of BB2.6 and a 3' region of XB3.6 were determined (Fig. 1A). The sequence of genomic DNA, containing 4008 bp in total, is presented in Fig. 2. The gene for GPAT was designated *ATSI* (for acyltransferase 1). Blots of some restriction digests of the *Arabidopsis* genome generated by combinations of *Eco* RI and *Eco* RV and also of *Sal* I and *Sty* I were probed with an *Arabidopsis* cDNA probe prepared as described above (data not shown). The numbers and sizes of the restriction fragments were those expected from the sequence of the genomic DNA, suggesting that the sequence of the gene represents the structure within the genome.

#### Isolation of *Arabidopsis* cDNA clones for a precursor to plastid-located GPAT

The *Arabidopsis* cDNA library was screened for hybridization with an *Arabidopsis* DNA probe prepared from the BB2.6 fragment of the genomic clone. After screening of 110 000 recombinant clones, one positive clone designated AR1 was obtained (Fig. 1B). Another set of 110 000 recombinant clones from the cDNA library was screened again for hybridization with an *Arabidopsis* cDNA probe prepared from the 1.3 kb *Sal* I-*Sty* I restriction fragment of AR1, and a positive clone designated AR2 was obtained (Fig. 1B). Sequence determination revealed that the cDNA inserts of the AR1 and AR2 clones were 1445 bp and 1725 bp in length, respectively, and they contained an overlapping sequence of

1424 bp (Fig. 1B). Consequently, we determined a cDNA sequence of 1746 bp which included a poly(A) sequence of 5 bp. An RNA blot of *Arabidopsis* poly(A)<sup>+</sup> RNA was probed with the *Arabidopsis* cDNA probe and a cross-reactive band of  $1.9 \times 10^3$  nucleotides in length was detected (Fig. 3). The primer extension of the RNA resulted in an extension product that extended 57 nucleotides beyond the first ATG codon (Fig. 4). This result suggests that the mRNA consists of 1783 nucleotides plus a poly(A) tail of ca. 120 nucleotides. Neither ATG codons nor possible intron sequences were found between the mapped 5' end and the first ATG codon, indicating that the cloned cDNA contains a full-length open-reading frame of 1377 bp. The nucleotide sequence flanking the putative initiation codon, 5'-GCTTTAATGACT-3', indicates that the reported consensus sequence for initiation of translation in plants, 5'-TAAACAATGGCT [10], was not included in the present case.

#### Deduced amino-acid sequence for a precursor to plastid-located GPAT from *A. thaliana*

The open-reading frame encodes a polypeptide of 459 amino acids. The amino-terminal leader sequence from 1 to 65 amino acid residues is rich

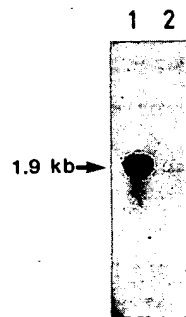


Fig. 3. RNA/DNA hybridization analysis of RNA from *Arabidopsis*. The RNA blot was probed with an *Arabidopsis* cDNA probe prepared from the 1.3 kb *Sal* I-*Sty* I restriction fragment of AR1. The size of the hybridized band was calibrated with RNA molecular weight marker I (Boehringer Mannheim, Germany) consisting of RNA markers 7.4, 5.3, 2.8, 1.9, 1.6, 1.0, 0.6, 0.4 and 0.3 kb long. 1, 5  $\mu$ g poly(A)<sup>+</sup> RNA; 2, 30  $\mu$ g total RNA.

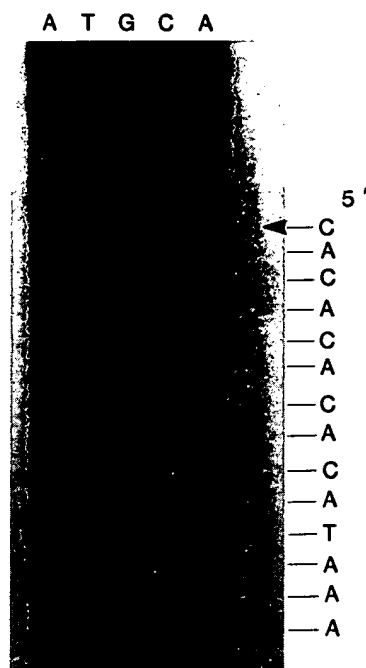


Fig. 4. Analysis of the 5' end of the RNA by primer extension. The primer extension reaction was performed using a procedure for RNA sequencing [6] and the products of transcription were analyzed on an 8% denaturing polyacrylamide gel. In parallel, DNA sequencing reactions were performed and products were used as size markers. For both reactions, the same oligonucleotide primer, namely, 5'-CGTGAGAGT-CATTAAAG-3', was used. Poly(A)<sup>+</sup> RNA (10 µg) and pXB3.6 were used as templates for the primer extension reaction and the sequencing reaction, respectively. The position of the detected transcript is shown by an arrow.

in hydroxy amino acids, serine and threonine, and small hydrophobic amino acids, alanine and valine, which together are indicative of a transit peptide [12]. The site for removal by cleavage of the leader sequence was not determined in the present study (see Discussion). The deduced amino-acid sequence of the polypeptide from *Arabidopsis* exhibited 68.8% and 60.4% homology to those of the plastid-located GPAT from squash (368 amino acids) [9] and a precursor to the plastid-located GPAT from pea (457 amino acids) [33].

#### Over-production of a GPAT fusion protein in *E. coli*

Figure 5 shows the results of SDS-PAGE of proteins from BL21 (DE3) cells transformed with

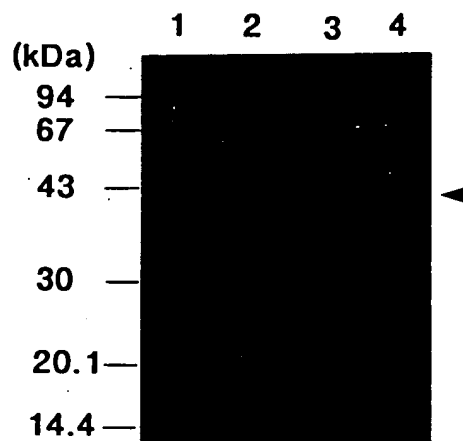


Fig. 5. The over-expression of glycerol-3-phosphate acyltransferase in *E. coli*. The proteins from *E. coli* cells transformed with pET-AR1 were analyzed by SDS-polyacrylamide gel electrophoresis in a 12.5% polyacrylamide gel. The 42-kDa protein is indicated by an arrow. 1, molecular mass markers; 2, total proteins from cells before addition of isopropyl- $\beta$ -thiogalactopyranoside (IPTG); 3, total proteins from cells 3 h after addition of IPTG; 4, a water-soluble protein fraction from cells 3 h after addition of IPTG.

pET-AR1. A 42 kDa polypeptide was synthesized upon addition to the bacterial culture of isopropyl- $\beta$ -thiogalactopyranoside (IPTG; Fig. 5, lane 3). The relative molecular mass of the protein was of the size anticipated for the fusion protein encoded by pET-AR1 ( $M_r$  42422). This polypeptide accounted for as much as 30% of the total water-soluble proteins after a further 3 h incubation (Fig. 5, lane 4). A substantial level of GPAT activity was detected in the soluble protein extracted from cells transformed with pET-AR1 but not from cells transformed with the vector plasmid, pET-3c, and the specific activity of GPAT was increased 6-fold after a 3 h incubation with IPTG (Table 1). These results suggest that the enzymatic activity is attributable to the water-soluble polypeptide and not to the membrane-bound glycerol-3-phosphate acyltransferases from *E. coli* [16]. Hence, we conclude that the *ATSI* gene is a structural gene that encodes a precursor to the plastid-located GPAT of *A. thaliana*.

Table 1. Over-expression of GPAT activity in soluble protein extracts from *E. coli* BL21 DE3).

Transformant	Activity (nmol/min)	Protein (mg)	Specific activity (nmol/min per mg protein)
BL21 (DE3)/pET-AR1			
- IPTG	1600	19	84
+ IPTG, then 3 h	14000	30	470
BL21 (DE3)/pET-3c			
- IPTG	0	18	0
+ IPTG, then 3 h	0	21	0

Cells were grown in 500 ml of broth until the optical density at 600 nm reached 0.5. Then half of the culture was kept for control measurements (- IPTG), and the other half was supplemented with 0.4 mM IPTG and incubated for another 3 h (+ IPTG, then 3 h). Water-soluble protein extracts were prepared and GPAT activity and protein contents were measured as described in the text.

## Discussion

Alignment between the genomic DNA and cDNA sequences revealed that the genomic DNA sequence contained 11 intron sequences of 1412 bp in total, each interrupting the nucleotide sequence of the open reading frame, and another intron sequence of 94 bp in the 3' untranslated sequence. All the intron sequences contained invariant GT and AG dinucleotides at the 5' and 3' splice sites, respectively. Sequences for intron-splicing sites were consistent with the consensus sequences for plant introns [7].

Sequences characteristic of expressed eukaryotic genes are found in the *ATSI* gene. However, the function of these sequences in the expression of the *ATSI* gene remains to be examined experimentally. No TATA box consensus sequences [10] are found within the expected distance, i.e.,  $32 \pm 7$  nucleotides upstream, from the mapped 5' end of the RNA. A TATA-like sequence at position -29, 5'-GGGATAAATAAAA-3', may play the role of a TATA box. The absence of a typical TATA sequence in housekeeping genes has occasionally been observed in animal and plant genes [2, 10]. The putative polyadenylation

signal, 5'-AATAAT-3', similar to the plant consensus, 5'-AATAAA-3' [11], is located 37 nucleotides upstream of the poly(A) site, almost within the expected distance ( $27 \pm 9$  nucleotides) [11].

In the present study, the 5' and 3' flanking sequences of the *ATSI* gene were also determined. However, an experimental examination of the possible significance of these sequences in the expression and regulation of the *ATSI* gene remains to be performed. A homology search indicated that a unique sequence, 5'-ATTCATTAT-3', is present in the 5' untranslated sequences of the *ATSI* gene and of one of the ACP genes from *Arabidopsis* [24]. However, the significance of this sequence in gene expression is unclear. The 3'-untranslated sequence contained an intron sequence, but it is not known whether the splicing of this intron serves any special function in the maturation and subsequent translation of the mRNA. The 3'-untranslated sequence of the mRNA contains two poly(A) tracts, one of which appears to have been annealed with oligo(dT) primers to cause priming for synthesis of cDNA, with resultant isolation of the AR1 clone. Finally, it should be noted that a one-base insertion into the cDNA sequence occurred in the 3'-untranslated region of the AR2 cDNA clone; the determined cDNA sequence contained a (T)<sub>15</sub> tract in the 3'-untranslated sequence, whereas the genomic DNA sequence from nucleotide number 3373 to 3386 contained a (T)<sub>14</sub> tract.

Figure 6 shows the alignment of amino-acid sequences of precursors to the plastid-located GPATs from *Arabidopsis*, pea [33] and squash [9]. Amino-acid sequences are conserved at least 7 regions (underlined in Fig. 6). The significance of these conserved sequences in the enzymatic activity and the interaction with substrates remains to be examined. In contrast to the conserved sequences, the amino-acid sequence from 1 to 90 of the precursor from *Arabidopsis* shows only 16.6% and 28.8% homology to the corresponding pre-sequences of the precursors from pea and squash, respectively. In particular, the amino-acid sequence from 61 to 90 of the precursor from *Arabidopsis* greatly differs from the cor-

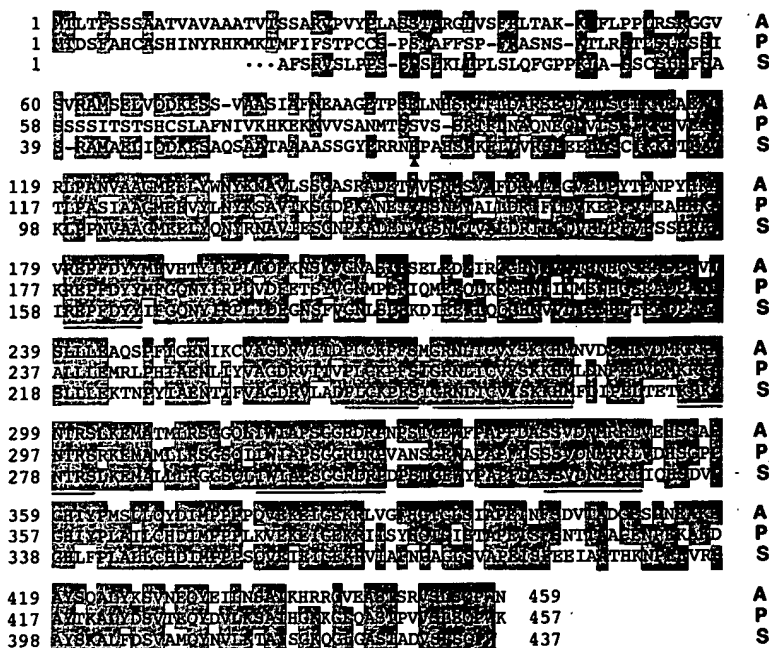


Fig. 6. The alignment of amino-acid sequences deduced from open-reading frames of cDNAs for precursors to plastid-located GPATs from *Arabidopsis*, pea [33] and squash [9]. Amino acids are given in one letter code, and gaps indicated by (-) are inserted to maximize similarity among sequences. Identical amino acids are shown in shadow boxes. A, P and S on the right column stand for *Arabidopsis*, pea and squash, respectively. Each amino-acid sequence is numbered separately on the left column. The deduced amino-acid sequence for a precursor to the GPAT from squash is truncated at the amino terminus [22, 33]. Amino-terminal residues of purified GPATs from squash [9] and pea [33] are indicated by an arrow.

responding sequence of the precursor from pea. Interestingly, however, the sequence of the precursor from *Arabidopsis* is similar to the aligned sequence of the precursor from squash. The amino-terminal residues of purified GPATs from squash [21] is  $^{70}\text{E}$ . The corresponding amino-terminal residue of mature GPAT from *Arabidopsis* is  $^{91}\text{E}$  (Fig. 6). If this is the amino terminus of the mature GPAT, the molecular mass of the mature polypeptide of the GPAT from *Arabidopsis* is calculated to be 41204 Da. However, we observed that the precursor from *Arabidopsis*, which had been over-expressed in transgenic tobacco and targeted into the chloroplasts, was estimated to be of 43 kDa by SDS-PAGE (Ishizaki, Nishida and Murata, unpublished result). Hence, it is likely that the processing site of the precursor to GPAT from *Arabidopsis* is located ca. 20 amino-acid residues upstream of the  $^{91}\text{E}$  residue. It is also likely that the purified GPATs from pea and

squash have been truncated off the amino-termini during purification.

In conclusion, we have cloned the structural gene from *A. thaliana* for a precursor to plastid-located GPAT. Cross-hybridization provides a useful approach for the isolation of this gene and the cDNA for plastid-located GPAT, at least from dicots. It will be of interest to determine whether the cloned DNA and cDNA can also serve as heterologous probes for detection of other genes for acyltransferases from higher plants and cyanobacteria. Cloning of the *ATSI* gene with flanking DNA sequences will allow the study of *cis* and *trans* regulatory elements of the gene. These studies and complementary studies of the gene for acyl-ACP hydrolase will allow evaluation of their regulatory role in the biosynthesis of glycerolipids in plant cells. As pointed out in our previous study [9], plastid-located GPAT from *A. thaliana* exhibits no homology to GPAT from



*Escherichia coli* [16]. Compilation of basic data, such as the sequences of the genes, RNAs and proteins for plant GPATs is a prerequisite for an understanding of the molecular diversity and evolution of GPATs in plants and animals.

### Acknowledgements

We thank Professors Y. Shimura and K. Okada for helpful suggestions and discussions. This work was supported in part by Grants-in-Aid for Co-operative Research (A) (No. 02304005), for Developmental Research (No. 01840027), for Scientific Research: on Priority Areas (No. 02261219, 03257103 and 04252102 and for Encouragement of Young Scientists (No. 01740402) from the Ministry of Education, Science and Culture of Japan.

### References

1. Ausubel FM, Brent R, Kingston RE, Moore DD, Seidman JG, Smith JA, Struhl K: Current Protocols in Molecular Biology, Wiley-Interscience, New York (1990).
2. Chen MJ, Shimada T, Moulton AD, Cline A, Humphries RK, Maizel J, Nienhuis AW: The functional human dihydrofolate reductase gene. *J Biol Chem* 259: 3933-3943 (1984).
3. Frentzen M: Comparison of certain properties of membrane bound and solubilized acyltransferase activities of plant microsomes. *Plant Sci* 69: 39-48 (1990).
4. Frentzen M, Heinz E, McKeon TA, Stumpf PK: Specificities and selectivities of glycerol-3-phosphate acyltransferase from pea and spinach chloroplasts. *Eur J Biochem* 129: 629-636 (1983).
5. Frentzen M, Neuburger M, Joyard J, Douce R: Intracellular localization and substrate specificities of the mitochondrial acyl-CoA: sn-glycerol-3-phosphate O-acyltransferase and acyl-CoA: 1-acyl-sn-glycerol-3-phosphate O-acyltransferase from potato tubers and pea leaves. *Eur J Biochem* 187: 395-402 (1990).
6. Hamlyn PH, Brownlee GG, Cheng CC, Gait MJ, Milstein C: Complete sequence of constant and 3' noncoding regions of an immunoglobulin mRNA using the dideoxynucleotide method of RNA sequencing. *Cell* 15: 1067-1075 (1978).
7. Hanley BA, Schuler MA: Plant intron sequences: evidence for distinct groups of introns. *Nucl Acids Res* 16: 7159-7176 (1988).
8. Hoey T, Warrior R, Manak J, Levine M: DNA-binding activities of the *Drosophila melanogaster* even-skipped protein are mediated by its homeo domain and influenced by protein context. *Mol Cell Biol* 8: 4598-4607 (1988).
9. Ishizaki O, Nishida I, Agata K, Eguchi G, Murata N: Cloning and nucleotide sequence of cDNA for the plastid glycerol-3-phosphate acyltransferase from squash. *FEBS Lett* 238: 424-430 (1988).
10. Joshi CP: An inspection of the domain between putative TATA box and translation start site in 79 plant genes. *Nucl Acids Res* 15: 6643-6653 (1987).
11. Joshi CP: Putative polyadenylation signals in nuclear genes of higher plants: a compilation and analysis. *Nucl Acids Res* 15: 9627-9640 (1987).
12. Keegstra K, Olsen LJ, Theg SM: Chloroplastic precursors and their transport across the envelope membranes. *Annu Rev Plant Physiol Plant Mol Biol* 40: 471-501 (1989).
13. Kunst L, Browse JB, Somerville C: Altered regulation of lipid biosynthesis in a mutant of *Arabidopsis* deficient in chloroplast glycerol-3-phosphate acyltransferase activity. *Proc Natl Acad Sci USA* 85: 4143-4147 (1988).
14. Laemmli UK: Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227: 680-685 (1970).
15. Löhden I, Frentzen M: Role of plastidial acyl-acyl carrier protein: glycerol 3-phosphate acyltransferase and acyl-acyl carrier protein hydrolase in channelling the acyl flux through the prokaryotic and eukaryotic pathway. *Planta* 176: 506-512 (1988).
16. Lightner VA, Bell RM, Mordrich P: The DNA sequences encoding *plsB* and *dkg* loci of *Escherichia coli*. *J Biol Chem* 258: 10856-10861 (1983).
17. Mizusawa S, Nishimura S, Seela F: Improvement of the dideoxy chain termination method of DNA sequencing by use of deoxy-7-deazaguanosine triphosphate in place of dGTP. *Nucl Acids Res* 14: 1319-1324 (1986).
18. Morre DJ: Membrane biogenesis. *Annu Rev Plant Physiol* 26: 441-481 (1975).
19. Murata N, Nishida I: Lipids in relation to chilling sensitivity of plants. In: Wang CY (ed) *Chilling Injury of Horticultural Crops*, pp. 181-199. CRC Press, Boca Raton, FL (1990).
20. Murata N, Ishizaki-Nishizawa O, Higashi S, Hayashi H, Tasaka Y, Nishida I: Genetically engineered alteration in the chilling sensitivity of plants. *Nature* 356: 710-713 (1992).
21. Nishida I, Frentzen M, Ishizaki O, Murata N: Purification of isomeric forms of acyl-[acyl-carrier-protein]: glycerol-3-phosphate acyltransferase from greening squash cotyledons. *Plant Cell Physiol* 28: 1071-1079 (1987).
22. Nishida I, Tasaka Y, Shiraishi H, Okada K, Shimura Y, Beppu T, Murata N: The genomic gene and cDNA for the plastidial glycerol-3-phosphate acyltransferase of *Arabidopsis*. In: Quinn PJ, Harwood JL (eds) *Plant Lipid Biochemistry, Structure and Utilization*, pp. 462-464. Portland Press, London (1990).

23. Ohlrogge JB: Biochemistry of plant acyl carrier protein. In: Stumpf PK (ed) *The Biochemistry of Plants*, vol. 9, pp. 137–157. Academic Press, Orlando, FL (1987).
24. Post-Beittenmiller MA, Hlousek-Radojcic A, Ohlrogge JB: DNA sequence of a genomic clone encoding an *Arabidopsis* acyl carrier protein (ACP). *Nucl Acids Res* 17: 1777 (1989).
25. Pruitt RE, Meyerowitz EM: Characterization of the genome of *Arabidopsis thaliana*. *J Mol Biol* 187: 169–183 (1986).
26. Rigby PWJ, Dieckmann M, Rhodes C, Berg P: Labeling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. *J Mol Biol* 113: 237–251 (1977).
27. Rosenberg AH, Lade BN, Chui D, Lin S-W, Dunn JJ, Studier FW: Vectors for selective expression of cloned DNA by T7 RNA polymerase. *Gene* 56: 125–135 (1987).
28. Roughan G, Slack R: Glycerolipid synthesis in leaves. *Trends Biochem Sci* 9: 383–386 (1984).
29. Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain terminating inhibitors. *Proc Natl Acad Sci USA* 74: 5463–5467 (1977).
30. Studier FW, Moffatt, BA: Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* 189: 113–130 (1986).
31. Stumpf PK: The biosynthesis of saturated fatty acids. In: Stumpf PK (ed) *The Biochemistry of Plants*, vol. 9, pp. 121–136. Academic Press, Orlando, FL (1987).
32. Soll J, Roughan PG: Acyl-acyl carrier protein pool sizes during steady-state fatty acid synthesis by isolated spinach chloroplasts. *FEBS Lett* 146: 189–192 (1982).
33. Weber S, Wolter F-P, Buck F, Frentzen M, Heinz E: Purification and cDNA sequencing of an oleate-selective acyl-ACP: *sn*-glycerol-3-phosphate acyltransferase from pea chloroplasts. *Plant Mol Biol* 17: 1067–1076 (1991).

**Communication****A Cerulenin Insensitive Short Chain 3-Ketoacyl-Acyl Carrier Protein Synthase in *Spinacia oleracea* Leaves**

Jan G. Jaworski\*, Richard C. Clough, and Susan R. Barnum

Chemistry Department (J.G.J.) and Botany Department (R.C.C., S.R.B.), Miami University, Oxford, Ohio 45056

**ABSTRACT**

A cerulenin insensitive 3-ketoacyl-acyl carrier protein synthase has been assayed in extracts of spinach (*Spinacia oleracea*) leaf. The enzyme was active in the 40 to 80% ammonium sulfate precipitate of whole leaf homogenates and catalyzed the synthesis of acetoacetyl-acyl carrier protein. This condensation reaction was five-fold faster than acetyl-CoA:acyl carrier protein transacylase, and the initial rates of acyl-acyl carrier protein synthesis were independent of the presence of cerulenin. In the presence of fatty acid synthase cofactors and 100 micromolar cerulenin, the principal fatty acid product of *de novo* synthesis was butyric and hexanoic acids. Using conformationally sensitive native polyacrylamide gel electrophoresis for separation, malonyl-, acetyl-, butyryl-, hexanoyl, and long chain acyl-acyl carrier proteins could be detected by immunoblotting and autoradiography. In the presence of 100 micromolar cerulenin, the accumulation of butyryl- and hexanoyl-acyl carrier protein was observed, with no detectable long chain acyl-acyl carrier proteins or fatty acids being produced. In the absence of cerulenin, the long chain acyl-acyl carrier proteins also accumulated.

It has been known for decades that fatty acid biosynthesis involves extending an acyl-ACP<sup>1</sup> stepwise by two carbons. In plants and bacteria, this synthesis is catalyzed by a type II fatty acid synthase, a readily dissociable group of enzymes which can be individually isolated and studied (1). One of the initial reactions leading to chain elongation is catalyzed by the enzyme 3-ketoacyl-ACP synthase, frequently referred to simply as the 'condensing enzyme.' The spinach fatty acid synthase was initially separated into its individual components by Shimakata and Stumpf (13, 14). In that study, they found two 3-ketoacyl-ACP synthases, designated I and II. 3-Ketoacyl-ACP synthase I was responsible for the majority of the condensations, using acyl-ACPs ranging from 2:0-ACP to 14:0-ACP. 3-Ketoacyl-ACP synthase II had its primary function in stearic acid synthesis, and thus was most active with 14:0-ACP and 16:0-ACP. A diagnostic difference between the two synthases was their sensitivity to the antibiotic cerulenin. 3-Ketoacyl-ACP synthase I was very sensitive and was inhibited 100% by as little as 10  $\mu$ M cerulenin. In contrast, 3-ketoacyl-ACP synthase II was inhibited only 50% by 50  $\mu$ M cerulenin. Thus, the characteristic sensitivity of fatty acid

biosynthesis to cerulenin can be attributed entirely to the inhibition of 3-ketoacyl-ACP synthase I.

Similarly, the 3-ketoacyl-ACP synthases found in *Escherichia coli*, are characteristically cerulenin sensitive (5). Recently, a third condensing enzyme has been reported in *E. coli*, which has been named acetoacetyl-ACP synthase (6). This 3-ketoacyl-ACP synthase is distinctly different from the other 3-ketoacyl-ACP synthases in several important respects. It is cerulenin insensitive, is specific for very short chain acyl-ACPs, and it prefers acetyl-CoA over acetyl-ACP (7). As a consequence of this last property, it would appear that this enzyme can bypass the much slower acetyl transacylase step and as a result, the acetyl transacylase would not be rate limiting in fatty acid biosynthesis.

In light of these findings in the *E. coli* system, we have reexamined the spinach system and report here the presence of a third 3-ketoacyl-ACP synthase which is completely insensitive to cerulenin, active only with short chain acyl thioesters, and appears to prefer acetyl-CoA to acetyl-ACP.

**MATERIALS AND METHODS****Materials**

Spinach (*Spinacia oleracea*) was purchased locally. Spinach ACP was the generous gift of J. Ohlrogge, Michigan State University. Malonyl-CoA was prepared by the method of Rutkoski and Jaworski (12). [<sup>14</sup>C]AcetylCoA was prepared from [<sup>14</sup>C]sodium acetate by preparing an acetyl-imidazole intermediate (3). Acyl-ACPs were prepared by the method of Cronan and Klages (4). All other chemicals were reagent grade or best available.

**Enzyme Preparation**

A crude spinach leaf extract was prepared by the method of Shimakata and Stumpf (14). The 40 to 80% ammonium sulfate precipitate was dissolved and dialyzed against 50 mM potassium phosphate (pH 8.0), 20% (v/v) glycerol, and 2 mM dithiothreitol. The protein concentration of this crude spinach leaf extract was 19 mg/mL as determined by the method of Bradford (2).

**Enzyme Assays**

The *in vitro* synthesis of acyl-ACPs was carried out under assay conditions for acetyl-CoA:ACP transacylase and fatty

<sup>1</sup>Abbreviation: ACP, acyl carrier protein.

acid synthase. For acetyl-CoA:ACP transacylase, the reaction mixture contained 100 mM Tris (pH 8.0), 0.5 mM 2-mercaptoethanol, 60  $\mu$ g of spinach ACP, 10  $\mu$ M [ $^{14}$ C]acetyl-CoA (56 mCi/mmol), and 15  $\mu$ L of crude spinach leaf extract (0.29 mg protein) in a total volume of 150  $\mu$ L. Malonyl-CoA:ACP transacylase was assayed the same as the acetyl transacylase, but 100  $\mu$ M [ $^{14}$ C]malonyl-CoA (10 mCi/mmol) replaced the acetyl-CoA. For fatty acid synthase assays, the reaction mixture contained in addition to the components of the acetyl transacylase reaction, 100  $\mu$ M malonyl-CoA, 1 mM NADH, 2 mM NADPH, and 100  $\mu$ M cerulenin when indicated. Cerulenin was maintained as a 1.0 mM stock solution (pH 4.0) at 4°C. Reactions were stopped by transferring an aliquot of the reaction mix at the designated interval to an equal volume of 10% (w/v) TCA, mixing, and storing on ice for 5 to 15 min. All samples were diluted with 1 mL of 5% TCA, centrifuged, and the supernatant discarded. Pellets were washed with an additional 1 mL of 5% TCA and then redissolved in 25  $\mu$ L of 50 mM Mes, approximate pH of 6.3. An aliquot of this solution was analyzed by liquid scintillation spectrometry or used immediately for subsequent analysis.

### Electrophoretic Separation and Analysis

Native PAGE (11), with 13% acrylamide and 1 mm thickness were run at 15 mA and at 15°C. Samples in 50 mM Mes were diluted with an equal volume of 20% (v/v) glycerol, 0.002% (w/v) bromophenol blue, and loaded onto gels directly. Proteins were transferred to nitrocellulose (8). For autoradiography, nitrocellulose was dried, and exposed to film for 48 h.

Immunoblots were carried out after proteins were transferred to nitrocellulose. Nitrocellulose was blocked with 5% (w/v) nonfat dry milk (Carnation Co., Los Angeles, CA) in TBST (10 mM Tris [pH 8.0], 150 mM NaCl, 0.05% [v/v] Tween 20) for 1 h. Spinach ACP I antibody was added to the blocking solution and incubated an additional 1 h. The antibody solution was poured off and the nitrocellulose was washed three times with TBST for 10 min each. Immunostaining was carried out using the anti-rabbit IgG alkaline phosphatase conjugate.

### HPLC

Acyl groups esterified to the ACP in the electrophoresis samples were analyzed on HPLC following mild alkaline hydrolysis. These samples were hydrolyzed by adjusting to approximately pH 9.5 and treating with 100 mM 2-mercaptoethanol for 1 h at 37°C (9). Samples were directly injected into an HPLC system equipped with ODS column, eluted by the solvent system indicated in Table I and the radioactive fractions collected. Retention time of each fatty acid was compared to the available standard.

## RESULTS AND DISCUSSION

### Initial Velocity Measurements

In *E. coli*, the acetoacetyl-ACP synthase was distinguished from the 3-ketoacyl-ACP synthases by its insensitivity to

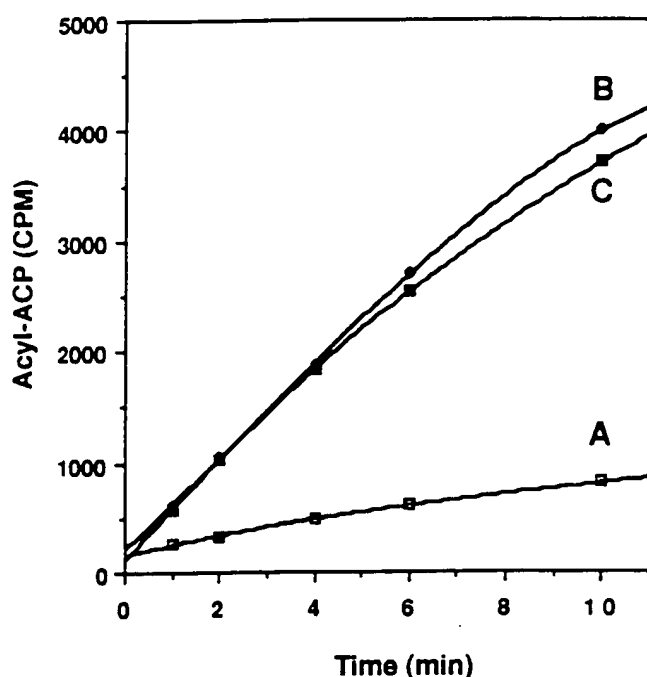
cerulenin (6). In addition, it preferred acetyl-CoA as a substrate and could be distinguished from acetyl transacylase by a 5- to 10-fold higher activity (7). Since Shimakata and Stumpf (15) had demonstrated that the acetyl transacylase was the slowest reaction step in fatty acid biosynthesis, we reasoned that if the acetyl transacylase step was bypassed by a more active 3-ketoacyl-ACP synthase, the overall rate of fatty acid biosynthesis should exceed the rate limiting acetyl transacylase step. The reactions used [ $^{14}$ C]acetyl-CoA as the only labeled substrate, in the absence of ATP, to ensure that any fatty acids synthesized would be labeled only once and thus have the same specific activity as the product of the acetyl transacylase reaction. Furthermore, the commonly employed methods for assaying fatty acid synthase were not used, in that these methods are dependent on measuring longer chain (10–18 carbons) fatty acids only. Since we were interested in measuring true initial velocities, we expected that a significant amount of the fatty acid synthase products would still be shorter chain fatty acids, that they would still be esterified to ACP, and they could be measured by acid precipitation. Acid precipitation was also the method used for the measurement of the acetyl transacylase, and thus would allow a direct comparison of the rates of the two reactions. The assay conditions for acetyl transacylase and fatty acid synthase differed only by supplementing the transacylase assay mixture with NADH, NADPH, and malonyl-CoA to allow for the 3-ketoacyl-ACP synthase and reductase reactions to occur. As illustrated in Figure 1, the initial velocity of the fatty acid synthase was 5-fold higher than the acetyl transacylase (fatty acid synthase: 11.7 pmol/min/mg protein *versus* acetyl transacylase 2.2 pmol/min/mg protein). These initial velocities were calculated from the linear portion of the progress curves, *i.e.* 1 to 4 min. The initial velocity data indicate that since the fatty acid synthase was much faster than the acetyl transacylase, the transacylase reaction was not rate-limiting and was bypassed by fatty acid synthase. Furthermore, it would indicate that acetyl-CoA rather than acetyl-ACP was the primer in these reactions.

To further examine if acetyl-CoA was preferred over acetyl-ACP as the primer for fatty acid synthesis, incubations containing 10  $\mu$ M [ $^{14}$ C]acetyl-CoA were supplemented with 10 and 20  $\mu$ M unlabeled acetyl-ACP. If there was equal or greater preference for acetyl-ACP over acetyl-CoA, the incorporation of label should have been decreased by 50% or more with equimolar acetyl-ACP. However, the incorporation of acetyl-CoA into fatty acids was diminished by only 25%, even with a twofold excess of acetyl-ACP. This decrease may even be attributed to dilution of label in the acetyl-CoA by the acetyl-ACP as a result of the reversible acetyl transacylase activity. Thus, this data was consistent with a strong preference for acetyl-CoA as the primer for fatty acid synthesis.

The effect of cerulenin on the rate of fatty acid synthase under these assay conditions was also determined (Fig. 1). The initial rate of fatty acid synthase was essentially unchanged by the addition of 100  $\mu$ M cerulenin. This concentration was 10-fold higher than the concentration needed to completely inhibit 3-ketoacyl-ACP synthase I (14), and indicated that a cerulenin insensitive 3-ketoacyl-ACP synthase was functioning in these reactions. Enzyme extracts were

**Table 1.** HPLC Data that Confirm the Identity of the Acyl Groups Attached to the ACPs

Acyl Group	Retention Time	Elution Solvent
	min	
Acetate	5.3	5 mM triethylamine (pH 5.0) adjusted with H <sub>3</sub> PO <sub>4</sub>
Malonate	6.1	5 mM triethylamine (pH 5.0) adjusted with H <sub>3</sub> PO <sub>4</sub>
Acetoacetate	6.5	5 mM triethylamine (pH 5.0) adjusted with H <sub>3</sub> PO <sub>4</sub>
Butyrate	6.3	50% MeOH in 10 mM H <sub>3</sub> PO <sub>4</sub>
Hexanoate	17.1	50% MeOH in 10 mM H <sub>3</sub> PO <sub>4</sub>

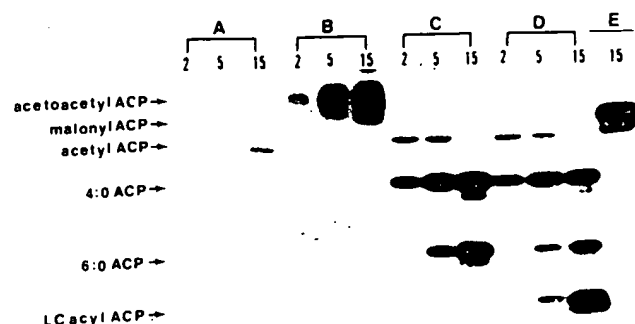


**Figure 1.** *In vitro* synthesis of acyl-ACP was measured under conditions for acetyl-CoA:ACP transacylase (curve A), fatty acid synthase activity (curve B), and fatty acid synthase activity with 100  $\mu$ M cerulenin (curve C) as described in "Materials and Methods." Reactions were stopped by transferring 25  $\mu$ L of the reaction mix at the designated interval to an equal volume of 10% (w/v) trichloroacetic acid (TCA), mixing, and storing on ice until the end of the experiment. Samples were prepared for liquid scintillation spectrometry as described in "Materials and Methods."

treated with cerulenin in several ways to avoid any artifacts due to stability or solubility of the antibiotic. Typically, a 1 mM stock at pH 4 stored 4°C was used (10). Freshly prepared solutions were also used as well as a freshly prepared ethanol solution (6), with no inhibition observed. Preincubation of the enzyme with cerulenin for as long as 30 min had no effect, as was the case with using lower cerulenin concentrations (10 and 50  $\mu$ M).

#### Acrylamide Gel Analysis of Products

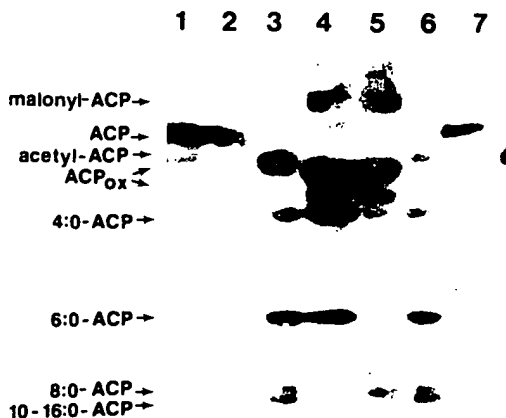
As the kinetic data were based on measuring acid precipitable radiolabeled products, it was necessary to confirm that these products were acyl-ACPs as well as to determine the identity of the acyl groups in the fatty acid synthase products. The native acrylamide gel system employed by Rock and Cronan (11) was used to successfully separate the acyl-ACPs



**Figure 2.** Autoradiogram of a native 13% PAGE of [<sup>14</sup>C]acyl-ACPs produced *in vitro* from spinach under the indicated reaction conditions. Enzyme assays were carried out and acyl-ACPs were prepared as described in "Materials and Methods." Each reaction was run for 2, 5, or 15 min, as indicated at the top of each lane. A, Acetyl-CoA:ACP transacylase—[<sup>14</sup>C] precursor was [1-<sup>14</sup>C]acetyl-CoA; B, fatty acid synthase with 100  $\mu$ M cerulenin, but omitting both NADPH and NADH; C, fatty acid synthase with 100  $\mu$ M cerulenin; D, fatty acid synthase without cerulenin present; E, malonyl-CoA:ACP transacylase (same as acetyl-CoA:ACP transacylase, but with [2-<sup>14</sup>C]malonyl-CoA as the substrate. Native gel was run at 15 mA and 15°C. Proteins were transferred to nitrocellulose, dried, and exposed to film for 48 h.

(Fig. 2). The acid precipitable product of the acetyl transacylase assay was exclusively [<sup>14</sup>C]acetyl-ACP, present in the expected low quantity. When the assay mixture was supplemented with only malonyl-CoA, a single product, acetoacetyl-ACP, was expected. There was only one radioactive band, tentatively identified as acetoacetyl-ACP, detected in these incubation products, migrating slightly behind the malonyl-ACP (compare lanes B and E, Fig. 2). These bands always lacked sharpness and we attribute this to instability of acetoacetate at pH 9.0. Supplementing this assay with NADH and NADPH resulted in production of saturated acyl-ACPs. When cerulenin was present, the major products were butyryl- and hexanoyl-ACP, with the precursor-product relationship apparent between them. In the absence of cerulenin, longer chain acyl-ACPs were also produced, as expected. Thus, these results were consistent with the presence of a cerulenin insensitive short chain 3-ketoacyl-ACP synthase, analogous to the acetoacetyl-ACP synthase found in *E. coli*.

The identities of the radiolabeled proteins as acyl-ACPs were determined by immunoblotting with spinach ACP antibody (Fig. 3) and HPLC analysis of the acyl groups (Table 1). The products of 15 min reactions (Fig. 3, lanes 4 and 5) were compared with standards and the bands observed on the autoradiogram. The radiolabeled bands corresponding to malonyl-, butyryl-, hexanoyl-, and long chain acyl-ACPs were



**Figure 3.** Immunoblot of a native 13% PAGE of ACPs and acyl-ACPs. Lanes 1, 2, 7, spinach ACP reduced with 5, 20, and 20 mM 2-mercaptoethanol, respectively. Lanes 3, 6, mixture of acetyl-, butyryl-, hexanoyl-, octanoyl-, decanoyl-, dodecanoyl-, and palmitoyl-ACP standards. Lane 4, products of 15 min fatty acid synthase reaction in the presence of 100  $\mu$ M cerulenin. Lane 5, products of 15 min fatty acid synthase reaction without cerulenin. Enzyme assays were carried out and acyl-ACPs were prepared as described in "Materials and Methods." Native gel was run at 15 mA and 15°C. Proteins were transferred to nitrocellulose, dried, and immunoblotted with an antibody to spinach ACP I.

immunoreactive with the ACP antibody. Several unlabeled immunoreactive bands were detected, including two uncharacterized forms of ACP. These two bands, which migrated just ahead of reduced ACP, disappeared upon reduction of the samples with 2-mercaptoethanol, and thus appeared to be oxidized forms of ACP. Because these bands were present in our incubation products, they obscured any acetyl-ACP band that might have been present. Acetyl-ACP (upper band in lanes 3 and 6) migrated at almost the same rate as one of these ACP bands. Since 2-mercaptoethanol catalyzed the rapid hydrolysis of the acyl-ACPs (data not shown), we could not add 2-mercaptoethanol to our samples (lanes 4 and 5) to examine for the presence of acetyl-ACP. Note in Figure 2, lanes C and D, that there was a minor radiolabeled protein, migrating between malonyl- and acetyl-ACP, which was consistently observed when NADH and NADPH was present during the incubation. This band is about where we expected to see reduced ACP (Fig. 3, lanes 1, 2, and 7), but there was no corresponding band visible in our samples on the immunoblots (Fig. 3, lanes 4 and 5). Thus, the labeled band observed on the autoradiogram presumably was not an acyl-ACP and was not further characterized. Finally, acetoacetyl-ACP could not be clearly identified on the immunoblots. The incubations which produced [ $^{14}$ C]acetoacetyl-ACP also contained high levels of unlabeled malonyl-ACP, which obscured the diffuse acetoacetyl-ACP band.

In addition to comparison with standard acyl-ACPs on PAGE, the identification of the acyl group esterified to ACP was also determined by HPLC. Each of the acyl-ACP samples was hydrolyzed, and the free acids analyzed directly by HPLC. Because of the wide range of hydrophobicity of the acyl groups, several different isocratic solvent systems (see Table

1) were used to analyze free acids. Samples containing only one or two radioactive acyl-ACPs were used in this analysis to eliminate ambiguity in the identification of the acyl-ACP. For each acyl-ACP, the retention of the corresponding free acid was compared with an authentic standard.

In summary, we have reported here that a third 3-ketoacyl-ACP synthase was identified in spinach. It had the distinctive characteristics of being cerulenin insensitive, specific for the synthesis of short chain fatty acids, *i.e.* C4 and C6, and apparently could use acetyl-CoA as a substrate. This enzyme appeared to be completely analogous to the *E. coli* acetoacetyl-ACP synthase reported by Jackowski and Rock (6) with an important distinction that *in vitro* products included C6 fatty acids. Thus, it would more appropriately be referred to as a short chain 3-ketoacyl-ACP synthase rather than an acetoacetyl-ACP synthase.

#### ACKNOWLEDGMENTS

The authors wish to thank John Ohlrogge for extensive and stimulating discussions during the course of this research and thank Susan Jackowski for her helpful suggestions.

#### LITERATURE CITED

- Bloch K, Vance D (1977) Control mechanisms in the synthesis of saturated fatty acids. *Annu Rev Biochem* 46: 263-298
- Bradford MM (1976) A rapid and sensitive method of quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72: 248-254
- Clough RC, Barnum SR, Jaworski JG (1989) Synthesis of radiolabeled acetyl-CoA from sodium acetate. *Anal Biochem* 176: 82-84
- Cronan JE Jr, Klages AL (1981) Chemical synthesis of acyl thioesters of acyl carrier protein with native structure. *Proc Natl Acad Sci USA* 78: 5440-5444
- D'Agnolo G, Rosenfeld IS, Awaya J, Omura S, Vagelos PR (1973) Inhibition of fatty acid synthesis by the antibiotic cerulenin. Specific inactivation of  $\beta$ -ketoacyl-acyl carrier protein synthetase. *Biochim Biophys Acta* 326: 155-166
- Jackowski S, Rock CO (1987) Acetoacetyl-acyl carrier protein synthase, a potential regulator of fatty acid biosynthesis in bacteria. *J Biol Chem* 262: 7927-7931
- Jackowski S, Murphy C, Cronan JE Jr, Rock CO (1988) Characterization of acetoacetyl-acyl carrier protein synthase and its inhibition by thiolactomycin. *FASEB J* 2: A1036
- Khyse-Andersen J (1984) Electrophoretic blotting of multiple gels: a simple apparatus without buffer tank for rapid transfer of proteins from polyacrylamide to nitrocellulose. *J Biochem Biophys Methods* 10: 203-209
- Lakin-Thomas PL, Brody S (1986) *Neurospora crassa* mitochondria contain two forms of a 4'-phosphopantetheine-modified protein. *J Biol Chem* 261: 4785-4788
- Omura S (1981) Cerulenin. *Methods Enzymol* 72: 520-532
- Rock CO, Cronan JE Jr (1981) Acyl-acyl carrier protein from *Escherichia coli*. *Methods Enzymol* 71: 341-351
- Rutkoski A, Jaworski JG (1978) An improved synthesis of malonyl-CoA. *Anal Biochem* 91: 370-373
- Shimakata T, Stumpf PK (1982) Fatty acid synthetase of *Spinacia oleracea* leaves. *Plant Physiol* 69: 1257-1262
- Shimakata T, Stumpf PK (1982) Isolation and function of spinach leaf  $\beta$ -ketoacyl-[acyl carrier protein] synthases. *Proc Natl Acad Sci USA* 79: 5808-5812
- Shimakata T, Stumpf PK (1982) The purification and function of acetyl coenzyme A:acyl carrier protein transacylase. *J Biol Chem* 258: 3592-3598

Short communication

## Analysis of two linked genes coding for the acyl carrier protein (ACP) from *Arabidopsis thaliana* (columbia)

Gayle Lamppa\* and Colleen Jacks<sup>1</sup>

Department of Molecular Genetics and Cell Biology, The University of Chicago, 920 East 58th St., Chicago, IL 60637, USA; <sup>1</sup>Current address: Department of Biology, Gustavus Adolphus College, St. Peter, MN 56082, USA

Received 20 September 1990; accepted in revised form 6 November 1990

**Key words:** acyl carrier protein, gene structure and expression, *Arabidopsis thaliana* (columbia)

### Abstract

Two linked genes, A1 and A2, coding for nearly identical isoforms of the acyl carrier protein (ACP) were isolated from an *Arabidopsis thaliana* (columbia) genomic library and sequenced. The amino acids deduced from the nucleotide sequence of the two genes indicate they encode distinct transit peptides, but the mature proteins are the same except for residue 79. Both genes are predicted to contain three introns in similar positions, although they differ in sequence and length. The introns interrupt regions coding for a) the transit peptide, b) the junction of the transit peptide and mature protein, and c) the highly conserved domain surrounding serine 38 to which the phosphopantetheine is attached. Primer extension analysis indicates that at least A1 is active in young plants.

The acyl carrier protein (ACP) is a key component in *de novo* fatty acid biosynthesis. It is a small acidic protein with a 4'-phosphopantetheine prosthetic group, attached through a serine located 38 residues from the N-terminus, to which growing fatty acid chains are covalently bound. In higher plants the major site of fatty acid biosynthesis appears to be the chloroplast (see for review Ohlrogge, 1987). Thus, ACP is synthesized as a precursor, imported into the organelle, and processed to its mature form. Recently, it was demonstrated that the chloroplast contains a holo-ACP synthase activity capable of transferring the prosthetic group from CoA, suggesting that production of the functional moiety occurs after translation into the organelle (Fernandez and Lamppa, 1990). At least two isoforms of ACP have been detected by SDS-PAGE and

peptide sequence analysis (Hoj and Svendsen, 1984; Ohlrogge and Kuo, 1985), but characterization of cDNA from *Brassica* (Safford *et al.*, 1988; Rose *et al.*, 1987), barley (Hansen and von Wettstein-Knowles, 1989) and spinach (Scherer and Knauf, 1987) indicates that other variants also exist in the cell.

An *Arabidopsis thaliana* (columbia) cosmid genomic library (Olszewski *et al.*, 1988) was screened with a *Brassica campestris* seed-specific cDNA clone (Rose *et al.*, 1987) coding for the ACP precursor. Cosmid DNA from a positive clone was digested with the enzymes shown in Fig. 1, and was predicted to contain ca. 18 kb of genomic DNA. Southern blot analysis using the *Brassica* cDNA insert (570 nt) as a probe showed that multiple bands hybridized in almost all of the restriction enzyme profiles. For example, in the

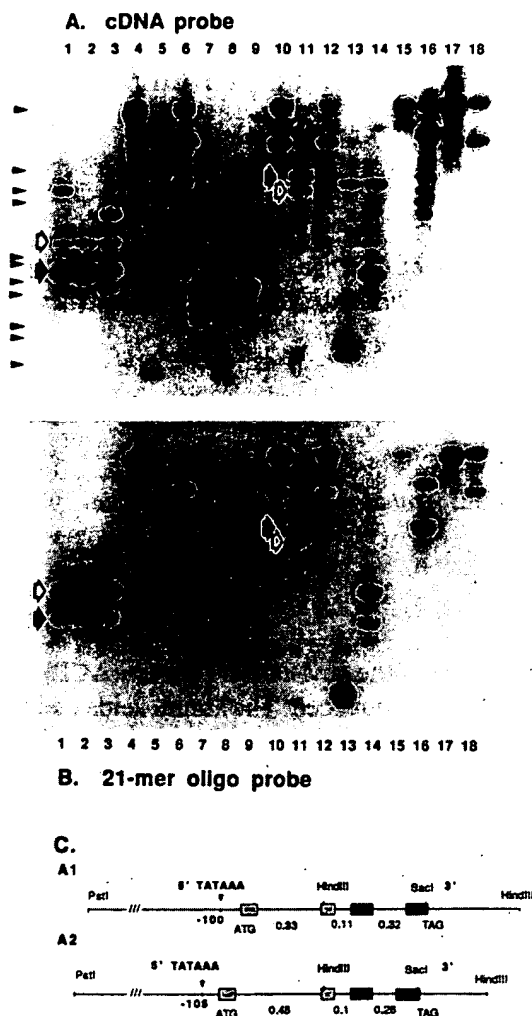


Fig. 1. Detection of two linked genes, A1 and A2, coding for ACP.

DNA was isolated from a cosmid clone that hybridized with a *Brassica* cDNA coding for ACP. The DNA was digested with the following restriction enzymes, as indicated, in lanes 1) *Hind* III, 2) *Hind* III/*Eco* RI, 3) *Hind* III/*Pst* I, 4) *Pst* I, 5) *Pst* I/*Eco* RI, 6) *Eco* RI, 7) *Eco* RI/*Sac* I, 8) *Eco* RI/*Pst* I, 9) *Kpn* I/*Sac* I/*Eco* RI, 10) *Kpn* I/*Sac* I, 11) *Pst* I/*Sac* I, 12) *Sac* I, 13) *Sac* I/*Hind* III, 14) *Kpn* I/*Hind* III, 15) *Kpn* I, 16) *Kpn* I/*Eco* RI, 17) *Kpn* I/*Pst* I, and 18) *Kpn* I/*Sac* I. Southern blot analysis was performed using either the radiolabeled *Brassica* cDNA insert (panel A), or a 21-mer oligonucleotide (panel B). The sequence of the degenerate oligonucleotide is CATNAC(A/G/T)AT(T/C)TC(T/C)ACNGT(A/G)TC. The cDNA was hybridized in  $3 \times \text{SSC}$  ( $1 \times \text{SSC} = 0.15 \text{ M NaCl}$ ,  $0.015 \text{ M Na-acetate}$ ) at  $48^\circ \text{C}$ , and washed in  $0.5 \times \text{SSC}$ ,  $0.1\%$  SDS at  $55^\circ \text{C}$ . The oligonucleotide was hybridized in  $3 \times \text{SSC}$  at  $42^\circ \text{C}$ , and washed in  $2 \times \text{SSC}$ ,  $0.1\%$  SDS at  $33^\circ \text{C}$ . In the *Hind* III and *Pst* I/*Sac* I profiles,

*Hind* III profile (Fig. 1A, lane 1) fragments of 3.9, 2.2 (open arrow) and 1.6 kb (closed arrow) hybridized with the cDNA insert that contains the entire coding sequence of the ACP precursor. The simplest pattern observed was after the *Pst* I/*Sac* I double digestion (Fig. 1A, lane 11), which showed two bands hybridizing with sizes of 3.7 (closed arrow) and 4.3 kb (open arrow). The complex hybridization profiles suggested the presence of several introns in a single ACP gene, or multiple ACP genes with internal restriction enzyme sites on the 18 kb genomic insert. To distinguish between these two possibilities, a degenerate 21-mer oligonucleotide was synthesized (see Fig. 1, legend), complementary to the sequence coding for residues DTVEIVM, which begin immediately carboxy to serine 38, and are highly conserved. In most profiles (Fig. 1B) the oligonucleotide hybridized to only two bands which were a subset of those that hybridized with the cDNA, strongly supporting the conclusion that the 18 kb insert contained two linked genes coding for ACP. The relative intensity of the hybridizing bands was about equal, in contrast to the pattern observed when the *Brassica* cDNA was used as a probe, suggesting that the *Brassica* sequence may be more distantly related to one of the two genes contained on these fragments. To substantiate this assessment, a partial restriction enzyme

the large arrows point to the fragments carrying part of either A1 (open) or A2 (closed), as assessed by their relative hybridization with the *Brassica* cDNA. The arrowheads in panel A (left side) indicate the position of lambda *Hind* III/*Eco* RI markers, with sizes (top to bottom) of 21.7, 5.15/5.0, 4.27, 3.48, 1.98, 1.9, 1.59, 1.37, 0.94, 0.83, 0.56, 0.14 kb. In panel C, a partial restriction enzyme map is shown with the orientation of each gene, and its basic structure. In addition to the *Hind* III-*Sac* I fragment, for gene A1, 990 nt were sequenced upstream of the internal *Hind* III site, and 185 nt 3' of the *Sac* I site. For gene A2, 760 nt were sequenced 5' of the internal *Hind* III site, as well as 370 nt 3' of the *Sac* I site. The exon boxes represent the transit peptide (dotted) and mature (black) regions of the ACP precursor. The numbers refer to the approximate size (kb) of each intron. The position of the TATA box relative to the start codon is indicated in nt. The distances between the *Pst* I and most distal *Hind* III site are ca. 5.6 kb and 5.7 kb for A1 and A2, respectively.



fragment map was deduced from the Southern analysis (Fig. 1C), and selected fragments were sequenced.

Two genes coding for ACP, called A1 and A2, were identified with their respective putative transit peptides (Fig. 2). They are 82% similar at the nucleotide level, and show 59% and 63% sequence relatedness (restricted to the mature protein coding sequence), respectively, to the *Brassica* cDNA initially used to identify them. The open reading frame for the mature protein encoded by A1 or A2 was based on the known amino acids of ACP from spinach (Kuo and Ohlrogge, 1984). Except for residue 79, which is phenylalanine (A1) or leucine (A2), the two mature proteins are identical at the amino acid level. One other *Arabidopsis thaliana* (*columbia*) ACP (Post-Beittenmiller *et al.*, 1989), referred to as the product of gene A3, differs significantly (23/83 residues), although a net charge of -9 is maintained.

The amino acid sequence of each transit peptide was predicted according to four criteria: a) the maximum similarity between the two sequences encoded by A1 and A2; b) the maximum similarity with the putative transit peptide deduced from A3 of *Arabidopsis* (Post-Beittenmiller *et al.*, 1989), and two *Brassica* cDNA clones (Rose *et al.*, 1987; Safford *et al.*, 1988); c) the fact that other reading frames are repeatedly interrupted by stop codons; and d) adherence to the GT/AG rule, and if possible the plant consensus sequences for exon/intron donor and acceptor sites (Brown, 1986). Starting with the initiator ATG, an open reading frame continues for 38 codons; another extends 5' from the beginning of the mature protein for 35 codons (Fig. 2). We predict an intron occurs between these two open reading frames. Although the start of the intron is difficult to predict, at the 3' end in both A1 and A2 a GCAG: G complies with the acceptor consensus sequence, and thus probably marks the start of the carboxy terminus of the transit peptide. As a consequence, the donor site must be in frame, and indicates that in A2 the sequence coding for the N-terminus of the transit peptide ends at the codon for leu 19, which is followed by the only GT

that does not interrupt a codon. There are three potential GT donor sites (see arrows, Fig. 2) in A1; the first coincides with the site found in A2, following leu 19. Hence, if the coding sequence is limited to the residues underlined in Figure 2, the transit peptides encoded by A1 and A2 are 80% related. Typically, they are rich in hydroxylated amino acids (serine and threonine) and the basic amino acid arginine.

The sequence analysis indicates that both A1 and A2 contain three introns in similar positions, although each intron differs in length between the two genes. Intron one interrupts the sequence coding for the transit peptide, as discussed above. Intron two (115 nt, A1; 100 nt, A2) occurs at the junction between the transit peptide and the mature protein, and may separate the codon for the first amino acid, alanine, from the remainder of the mature protein. Intron three (324 nt, A1; 264 nt, A2) begins 3 codons downstream of the special serine 38, and contains an internal block of ca. 160 nt (starting at nucleotide 1329 in A1, and 1139, A2) that is ca. 90% similar between the two genes. Introns two and three occur in the same positions in A1, A2 and A3, as well as in two *Brassica* genes (deSilva *et al.*, 1990).

A TATA box (TATAAA) is found 100 nt 5' of the ATG in gene A1 and 105 nt upstream of A2. The ATG is situated within a sequence motif (TCTATGGC) that follows the plant consensus sequence (ACAATGGC) for translation initiation (Joshi, 1987a). At the 3' end, neither gene contains an AATAAA, the consensus sequence for polyadenylation found in transcripts from animal cells, within 150 nt of their stop codons. However, gene A2 has an AATAAT and an AATGAA present 232 nt and 259 nt, respectively, 3' of its translation stop, similar to that observed in other ACP encoding genes (Safford *et al.*, 1988) and plant genes in general (Joshi, 1987b).

To determine if A1, and possibly A2, is actively transcribed, an oligonucleotide was synthesized to use as a primer in reverse transcription reactions. The oligonucleotide (5'-GCGAGGAC-GAGCCTTGAAGGG-3') is the reverse complement of the codons for amino acids 11-17 in the

<b>A1</b>	CATAGTAAATTTGATGCGTACAAAAATGAATGTAATTTGATTACAGACATAA	<b>A2</b>	TTTTTCCGAATTTGTTATTCTCTACCTGCATCTGCCAAATCACCACC
51	CAGAAATGACAACGTAACCTTTATAACACAATAAATAGTAATCAAAAT	51	TCACCTGACTTGCTGTCTGAATTTCTCATTGGAAATCTCAGAAAAGTTT
101	ATACAAATGCAACGAATAAATTTCTGAAATTTGTTTCGTATAAAAAAA	101	TTTTTTTTTTTTTTTGGTTAGGTTAAAAATACTATAAAATTAAGTCT
151	TTTAAACAGTAAACATAATTAAATGTAAAAATCTTTGTGATTTTTTT	151	CCCACTTGGATTCTTCACGTGTCTCACATCTCTTTGTCTTCTCCGCTC
201	TTTGTAAGAAAGACGTAGTAGAATTTACAAATTAAGGCATAAAATGTAT	201	CTCCGATCTCACTCCGATCTCTCTACGATTCATTCTCTATGGCTTCCAT
251	AACAGATAAAACCATATAAGTTTCTTAATATTTTCTTAATTTTTGCCT	251	eAlaThrSerAlaSerThrSerLeuGlnAlaArgProArgGlnLeuValS
301	TTTGTTTCTTGTCTAGGTTTATTTTCCACCGTCACTGCCAAGCACCAC	301	erLeuPheSerSerSerProArgIleSerHisLeuPheLeuThrGluPhe
351	TCCACCTAGCTTCTCTGATTCTGTGTCTTCCGCTTCAACTTTCATTGGAA	351	LeuSer
401	TCTCAGATTTTTTATTTTTTGTGTTTGGTTTGGTGGAGTTTGTGA	401	CTTAGTTGTTCTTCTTCTTCTCACTGTAGATCTAATTCGCAGATTCGT
451	CTATAAAACCTCTCCCACTTGGTTCTTCACTCTCACTGTTTCTCATCTCT	451	TTTCTCATCTCTATATTGACTGTTGAAATGCTCTGTAATTCATTTTTTG
501	TCGTCTTCTCCGCTCTCTCAATCTCACTCCGATCTCTCTACGATTCATT	501	TCCTGTAATTTTTTTCATTGAGGTAAGTTAGTTTTCTGTTTTCTGAATTT
551	MetAlaSerIleAlaAlaSerAlaSerIleSerLeuGlnAlaAr	551	TAGGCTCTGTTGAACCTATTTTACTTAACCGGTGCAATCACCAGAAAAA
601	gProArgGlnLeuValSerLeuPheProArgIleSerHisLeuValLeuL	601	TTATCCTATTCTCTTGAATTTGATTGTTTTATATCTGAATTTGAATA
651	euValLeuLeu	651	GATTCGAACCTCTGGTAAGAGCATGCGGGTTACATAGATTTTAGTTGCT
701	TGTAGATCTAATTTGCAGAATCGTTTTCTGAGTTCTAGTTGATTGTTT	701	TTGTTTCTCTTTTGGTTTGGTTTTCAGTGCAAAATATATTGCTTTCATT
751	AGGTGGCTCTATGATTTAGTTTTCTGGAACCTTGGCTCTGATAGCTGAT	751	TGCATTGGAGTTTAGTTTTGTACTGATGAAAGTTATATTGCAATATGA
801	GAATCTGTTTACTATTCTCCGTTGCTAATCACCAGAAATATTGTTGATT	801	TrpGlnValIleGlyAlaLysGlnValLysSerPheSerTyrGlySerAr
851	TCTCGTCGAATTGAATCACATAGATTTAGTTGCTTCTGTTAACTTCTC	851	gSerAsnLeuSerPheAsnArgArgGlnLeuProThrArgLeuThrValT
901	TCGTAGATTGTTTCGTTATTTTGGTTTGTGCTTTGCATTACGAGTTTAA	901	yrCysAla
951	TrpGlnAlaIleAlaAlaSerGlnValLysSerPheSer	951	TTACTTTTGATCTCTCGAGCCAATTCATCATTACCATTGTTGATTGTTG
1001	AsnGlyArgArgSerSerLeuSerPheAsnLeuArgGlnLeuProThrAr	1001	AlaLysProGluThrValAspLysValCysAlaValValArg
1051	gLeuThrValSerCysAla	1051	LysGlnLeuSerLeuLysGluAlaAspGluIleThrAlaAlaThrLysPh
1101	TGTTCAACAATGTGATTAGGCTCTTTGATACGTTTGGGTGATTGGTTGAT	1101	eAlaAlaLeuGlyAlaAspSerLeuAspThr
1151	AlaLysProGluThrV	1151	TGCTGCACTTGGTGTCTGATTCCTTGTATACGGTAACCTACTTCTCATGTC
1201	alAspLysValCysAlaValValArgLysGlnLeuSerLeuLysGluAla	1201	TAATAATTGGATTCAACTTTTGAATGGGATTGTTTCATTGGACTTATTGA
1251	AspGluIleThrAlaAlaThrLysPheAlaAlaLeuGlyAlaAspSerLe	1251	GTCGAATGTTTAGATTGCAAGCTTGTGAAGGATCGCTGCTTCTCCTTAT
1301	uAspThr	1301	ACACACATCTCTAAACATGTTACTGAGATACACATACTCATGATTCTTGT
1351	CTTTTGAATGGGATTGTTTCATTGGACTTATTGATGCTTAATCAATGCCTC	1351	ValG1
1401	TTAAAAATTGTATCTGCTTTCAAGCTCATATTGTGCAATGTTTATAGATT	1401	uIleValMetGlyLeuGluGluGluPheGlyIleGluMetAlaGluGluL
1451	GCAAGTCTTGTAAGGGTCTGCTGTCTCTTATACGCAGCTCTTTAGTA	1451	ysAlaGlnSerIleAlaThrValGluGlnAlaAlaLeuIleGluGlu
1501	TTGTTACTGATACAGGCAAAATATATAGGACATACTCAACACATACTTCA	1501	LeuLeuLeuGluLysAlaLys
1551	CAAGCAATGCATTATCAAAATTAATTCAAAAATGGATTGCAACGTGTGCA	1551	CTCTTGTGAAAGGCCAAGTAGAAAAATCCTTTACTACTTAGCAAAAAA
1601	CATACTCATGACTCGCGCTTTTGGATTTCAGGTGGAGATTGTGATGGGAC	1601	CGGAAAAAACCAAAACCAACCACTATCTTCTTATTGGTTTTGGTTAG
1651	euGluGluGluPheGlyIleGluMetAlaGluGluLysAlaGlnSerIle	1651	CTAGAGAGCAATTCGTCGTCGTTGAGATTTAATGGCTATTTTCGGATA
1701	AlaThrValGluGlnAlaAlaAlaLeuIleGluGluLeuLeuPheGluLy	1701	ATTATTACAAGGCTGTATTTTAAATATTACTTTCTGTTCTGAAACATTTT
1751	sAlaLys	1751	AAATAATATAATTATGGTTGATGACTTCTTAATGAAAGGATTGAGTTTC
1801	ACCCAAACCACATCTTATTGTTTCTGTAGTAGAGAGCAATTGTGTCT	1801	TTGACATTTTTTGTCTTTTATAATCTAGAAGTTTGTGTCGGAAGAT
1851	GTTAAAGATTTTATGTTATTTTGGGAAATTTATACAAGGCTTGTATCTA	1851	TCCAAATTCAGGGGAAGGGGCATAAA 1876
1901	CTTTTATCTTCTCTTAAACCATTTT 1927		

Fig. 2. Nucleotide sequences of ACP genes A1 (left) and A2 (right). The predicted open reading frame (see text) of each gene is given in the three-letter amino acid code above the nucleotide sequence determined by the dideoxy chain termination method (Sanger *et al.*, 1977). A putative transit peptide sequence is underlined; the small and medium-sized arrows below the line indicate the possible donor and acceptor sites, respectively, of intron one (see text). The large arrow 5' of the initiator ATG indicates the transcription start site based on primer extension analysis. The TATA box and potential polyadenylation signals are underlined. The asterisk denotes the start of a highly conserved 5' region between A1 and A2.

transit peptide (SLQARPR) of gene A1, and differs in only three positions (underlined) from the analogous sequence of A2. The oligonucleotide sequence is not found in A3 (Post-Beitenmiller *et al.*, 1989). RNA was isolated from young *Arabidopsis* plants grown on MS/sucrose medium and hybridized ( $0.2 \text{ mg mL}^{-1}$  at  $21^\circ\text{C}$ ) with the radiolabeled oligonucleotide ( $2 \times 10^6 \text{ CPM pmol}^{-1}$ ). Following primer extension with M-MLV reverse transcriptase, the hybrids were digested with RNase as described (Sambrook *et al.*, 1989), and immediately analyzed by denaturing gel electrophoresis. A 123 nt RNase-resistant band was found (not shown). These results indicate that transcription begins 72 nt upstream of the ATG codon in A1, and possibly A2 as well, although a stable hybrid between the primer and A2 transcripts may not be favored. The transcription start site places the TATA box of gene A1 28 nt upstream (and potentially 34 nt upstream for A2). The 5' nontranslated regions of A1 and A2 are 96% related and rich in pyrimidines.

The 5' flanking regions of genes A1 and A2 both contain a highly conserved domain beginning ca. 95 nt from the transcription start with the sequence TCATTGGAATCTCAGA, followed by at least 20 thymidines, occasionally punctuated by a purine. This sequence is not found 5' of the A3 gene, which codes for a different isoform of ACP (Post-Beitenmiller *et al.*, 1989). It is interesting to speculate that the sequence is important in selectively regulating the expression of A1 and A2. This question will be most fruitfully addressed by the analysis of transgenic plants carrying a reporter gene under the control of either the A1 or A2 promoter.

### Acknowledgements

We wish to thank Dr Vic Knauf (Calgene, Inc., Davis, CA) for providing us with the *Brassica* cDNA, Dr Neil Olszewski (Dept. of Botany, Univ. of Minnesota) for the *Arabidopsis* genomic library, and Arthur Milgram for helping with subcloning of restriction enzyme fragments. Roben

Buell and Jia Zhao provided valuable technical assistance. This research was supported by the United States Dept. of Agriculture (grant no. 89-37261-4471).

### References

- Brown JW: A catalogue of splice junction and putative branch point sequences from plant introns. *Nucleic Acids Res* 14: 9549-9559 (1986).
- deSilva J, Loader NM, Jarman C, Windust JH, Hughes SG, Safford R: The isolation and sequence analysis of two seed-expressed acyl carrier protein genes from *Brassica napus*. *Plant Mol Biol* 14: 537-548 (1990).
- Fernandez MD, Lamppa GK: Acyl carrier protein (ACP) import into chloroplasts does not require the phosphopantetheine: evidence for a chloroplast Holo-ACP synthase. *Plant Cell* 2: 195-206 (1990).
- Hansen L, von Wettstein-Knowles PV: Acyl carrier proteins of barley seedling leaves and caryopses. In: Biacs PA, Gruiz K, Kremmer T (eds) *Biological Role of Plant Lipids*, pp. 367-370. Akademial Kiado/Plenum Publish. Corp., New York (1989).
- Hoj PB, Svenden I: Barley chloroplasts contain two acyl carrier proteins coded for by different genes. *Carlsberg Res Commun* 49: 483-492 (1984).
- Joshi CP: An inspection of the domain between putative TATA box and translation start site in 79 plant genes. *Nucleic Acids Res* 15: 6643-6653 (1987a).
- Joshi CP: Putative polyadenylation signal in nuclear genes of higher plants: a compilation and analysis. *Nucleic Acids Res* 15: 9627-9641 (1987b).
- Kuo TM, Ohlrogge JB: The primary structure of spinach acyl carrier protein. *Arch Biochem Biophys* 234: 290-296 (1984).
- Ohlrogge JB, Kuo TM: Plants have isoforms for acyl carrier protein that are expressed differently in different tissues. *J Biol Chem* 260: 8032-8037 (1985).
- Ohlrogge JB: Biochemistry of plant acyl carrier proteins. In: Stumpf PK (ed) *The Biochemistry of Plants*. Vol. 9. *Lipids: Structure and Function*, pp. 137-157 (1987).
- Olszewski NE, Martin FB, Ausubel FM: Specialized binary vector for plant transformation: expression of the *Arabidopsis thaliana* AHAS gene in *Nicotiana tabacum*. *Nucleic Acids Res* 16: 10765-10782 (1988).
- Post-Beitenmiller MA, Hlousek-Radojcic A, Ohlrogge JB: DNA sequence of a genomic clone encoding an *Arabidopsis* acyl carrier protein (ACP). *Nucleic Acids Res* 17: 1777 (1989).
- Rose RE, DeJesus CE, Moylan SL, Ridge NP, Scherer DE, Knauf VC: The nucleotide sequence of a cDNA clone encoding acyl carrier protein (ACP) from *Brassica campestris* seeds. *Nucleic Acids Res* 17: 719 (1987).

Safford R, Windust HC, Lucas C, Silva JD, James CM, Hellyer A, Smith CG, Slabas AR, Hughes SG: Plastid-localised seed acyl-carrier protein of *Brassica napus* is encoded by a distinct, nuclear multigene family. Eur J Biochem 174: 287-290 (1988).

Sambrook J, Fritsch E, Maniatis R: Molecular Cloning, a Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989).

Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74: 5463-5467 (1977).

Scherer DE, Knauf VC: Isolation of a cDNA clone for the acyl carrier protein-I spinach. Plant Mol Biol 9: 127-134 (1987).

was digested with Eco RI, fractionated on 0.7% agarose, blotted onto Hybond-N (Amersham), and hybridized as described [J. Sambrook, A. Frisch, T. Maniatis, *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1989)].

16. Protoplasts from *N. tabacum* SF S-Breznovits, L. Marton, F. Joo, (1975)] were isolated [J. Drap Armitage, R. Walden, *Plant Gen and Gene Expression: A La* (Blackwell, Oxford, 1988)] and *Agrobacterium* that contained pP described [A. Depicker *et al.*, *N.* 201, 477 (1985)]. Protoplasts were cultured in the presence of hygromycin (15  $\mu$ g/ml) (Boehringer Mannheim) and Claforan (500  $\mu$ g/ml) (Hoechst,

Frankfurt, Germany) with either cytokinin (kinetin, 0.2 mg/l) (Sigma) and auxin (1-naphthaleneacetic acid, NAA, 1.0 mg/l) (Sigma) or the same media lacking auxin. Calli growing in the absence of auxin were cultured further to form plants.

### Claim 67 Fatty acid desaturase

manuscript.

29 April 1992; accepted 1 September 1992

## Map-Based Cloning of a Gene Controlling Omega-3 Fatty Acid Desaturation in *Arabidopsis*

Vincent Arondel, Bertrand Lemieux, Inhwan Hwang, Sue Gibson, Howard M. Goodman, Chris R. Somerville\*

A gene from the flowering plant *Arabidopsis thaliana* that encodes an omega-3 desaturase was cloned on the basis of the genetic map position of a mutation affecting membrane and storage lipid fatty acid composition. Yeast artificial chromosomes covering the genetic locus were identified and used to probe a seed complementary DNA library. A complementary DNA clone for the desaturase was identified and introduced into roots of both wild-type and mutant plants by Ti plasmid-mediated transformation. Transgenic tissues of both mutant and wild-type plants had significantly increased amounts of the fatty acid produced by this desaturase.

The small crucifer *Arabidopsis thaliana* (L.) is suitable for the application of map-based cloning methods because it has a small nuclear genome that is almost devoid of interspersed, highly repetitive DNA (1). The five chromosomes have a total DNA content of about 70,000 kb (1). The average distance from any gene to the nearest restriction fragment length polymorphism (RFLP) marker (2, 3) is about 225 kb, and several yeast artificial chromosome (YAC) genomic libraries are available for *Arabidopsis* (4, 5). Here, we used these resources to isolate a gene from *Brassica napus* that complements a mutant of *Arabidopsis* deficient in omega-3 unsaturated fatty acids.

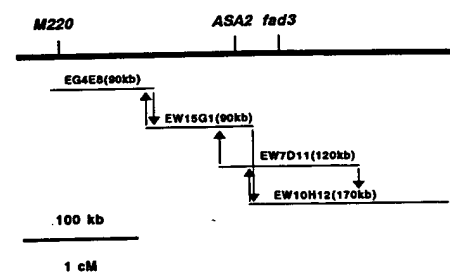
Fatty acid desaturases catalyze the  $O_2$  and electron donor-dependent insertion of double bonds into fatty acids. Because the amount of membrane lipid unsaturation affects the physical properties of membranes, the desaturases are thought to affect the ability of plants to survive extreme temperatures (6). Also, the nutritional quality of

edible plant oils is largely determined by the composition of storage triacylglycerols. Except for the stearyl-ACP desaturase, plant desaturases are thought to be membrane proteins and have been difficult to characterize by conventional biochemical methods (7). Information about the number and properties of the various desaturases in *Arabidopsis* has been obtained by the isolation of an extensive collection of mutants with altered membrane and storage lipid unsaturation (8). One of the mutations, designated *fad3*, resulted in reduced accumulation of linolenic acid ( $18:3^{\omega 3,6,9}$ ), and a corresponding increase in the amount of linoleic acid ( $18:2^{\omega 6,9}$ ), in extrachloroplast membrane and storage lipids (9). These metabolic effects suggested that the *fad3* locus encoded an  $\omega 3$  linoleate desaturase.

The *fad3* locus was genetically mapped by scoring the fatty acid composition of 137 progeny (F2) of a cross between plants of the *fad3* mutant line BL1 (Landsberg race) (9) and wild type (Niederzenn race). Because the *fad3* phenotype is only weakly evident in chlorophyllous tissues, but is strongly expressed in root or seed tissue, the fatty acid phenotype of each F2 plant was scored by gas chromatographic analysis of the fatty acid composition of each of ten seeds obtained by self-fertilization of F2 plants. The RFLP genotype of F2 progeny was determined by analysis of DNA prepa-

rations from F2 plants and F3 families (2, 3). The *fad3* mutation mapped on chromosome 2 adjacent to the RFLP markers M220 (2) and ASA2 (10) (Fig. 1). When used to screen YAC libraries, M220 hybridized with YAC EG4E8 (4) and ASA2 hybridized with YACs EW7D11 and EW15G1 (5). The YACs were ordered by analysis of hybridization of end-specific probes on Southern (DNA) blots of the YAC clones. The resultant, approximately 340 kb, contig of four YACs extended a minimum of 170 kb to the right of ASA2 (Fig. 1). As the maximum distance between M220 and ASA2 was 180 kb (the sum of the inserts in EG4E8 and EW15G1), we estimated that 1 centimorgan (cM) was equivalent to less than an average value of 105 kb in this region. As *fad3* was 0.4 cM to the right of ASA2, we estimated that the YAC contig extended far enough to the right to include *fad3*.

Stearyl-ACP desaturase, an enzyme that catalyzes a chemically equivalent reaction, is encoded by a moderately abundant mRNA in developing seeds of oil-accumulating plants such as *Ricinus communis* (11). Therefore, we used the YACs to isolate moderately abundant cDNA clones. DNA from one YAC, EW7D11, was isolated from a low-melting agarose-clamped homogeneous electric field (CHEF) gel (12) and used to probe a  $\lambda$ gt11 cDNA library made from developing seeds of the closely related crucifer *Brassica napus*. The *B. napus* library was used because we did not have a suitable library from developing seeds of *Arabidopsis*, and genes from *B. napus* are highly homologous to the corresponding *Arabidopsis* genes (13). Of 31 positive plaques among  $2 \times 10^5$  screened, 17 cross-hybridized at high stringency and, therefore, appeared to be derivatives of the same gene. None of the other clones were highly represented among the 31 positive clones. The largest insert (1.4 kb) representing the abundant transcript was



**Fig. 1.** Genetic map of the region of chromosome 2 (thick line) that contains the *fad3* locus. The YACs that correspond to this region of the genome are shown below. The sizes of the inserts (in parentheses), were determined by pulsed-field gel electrophoresis (12). Arrows indicate regions of overlap demonstrated by hybridization of end-specific probes. Scales of genetic distance (centimorgans, cM) and kilobases (kb) are indicated.

V. Arondel, S. Gibson, C. R. Somerville, MSU-DOE Plant Research Laboratory, Michigan State University, East Lansing, MI 48824-1312.  
B. Lemieux, Department of Biology, York University, Toronto, Canada M3J 1P3.  
I. Hwang and H. M. Goodman, Department of Genetics, Harvard Medical School, and Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114.

\*To whom correspondence should be addressed.

**Table 1.** Fatty acid composition of transgenic roots. The transgenic roots resulting from infection of the *fad3* mutant or wild type with *A. tumefaciens* R1000 carrying only the vector (pBI121) or the vector plus cDNA (pTiDES3) were grown in the presence of kanamycin (50 µg/ml) for 3 weeks to identify roots that had been co-transformed with one of these plasmids and 200 µg of cefotaxime to prevent growth of bacteria. The fatty acid composition of samples of roots (about 30 mg per sample) was determined as described (23). Abbreviations: 16:0, palmitic acid; 16:1, palmitoleic acid; 18:0, stearic acid; and 18:1, oleic acid; 18:2, linoleic acid; 18:3, linolenic acid. The values presented are the mean  $\pm$  SD ( $n = 12$ ).

Genotype	Fatty acid					
	16:0	16:1	18:0	18:1	18:2	18:3
	Percentage of total fatty acids					
Wild type	22.0 $\pm$ 2.9	2.5 $\pm$ 0.7	2.3 $\pm$ 1.9	3.8 $\pm$ 1.3	37.3 $\pm$ 3.7	31.9 $\pm$ 4.5
pBI121						
<i>fad3</i>	21.2 $\pm$ 1.6	1.6 $\pm$ 0.8	2.3 $\pm$ 1.5	5.9 $\pm$ 2.6	62.2 $\pm$ 5.9	6.7 $\pm$ 0.7
pBI121						
<i>fad3</i>	21.3 $\pm$ 2.3	1.5 $\pm$ 0.2	1.6 $\pm$ 0.4	9.1 $\pm$ 2.0	24.4 $\pm$ 14.9	42.1 $\pm$ 15.5
pTiDES3						
Wild type	21.1 $\pm$ 0.9	2.0 $\pm$ 0.1	1.9 $\pm$ 0.2	7.7 $\pm$ 2.0	15.7 $\pm$ 11.7	51.3 $\pm$ 10.9
pTiDES3						

subcloned into pBluescript (Stratagene) to produce plasmid pBNDES3.

In order to test if the cDNA in pBNDES3 was encoded by the *B. napus* equivalent of the *fad3* locus, a genetic complementation test was conducted. Because the *fad3* mutation has a large effect on the fatty acid composition of *Arabidopsis* roots (9), we exploited the fact that large numbers of rooty tumors can be rapidly produced by infection with *Agrobacterium tumefaciens* R1000, which carries an Ri plasmid from *Agrobacterium rhizogenes* instead of a Ti plasmid (14). The cDNA insert from pBNDES3 was inserted into the binary Ti vector pBI121 (Clontech Laboratories, Palo Alto, California) under transcriptional control of the constitutive cauliflower mosaic virus (CaMV) 35S promoter, to produce plasmid pTiDES3. After electroporation of pTiDES3 into *A. tumefaciens* R1000, the bac-

teria were used to induce rooty tumors on stem explants from the *fad3* mutant and wild-type *Arabidopsis* plants (15). More than 50% of the rooty tumors produced in this way contain the binary Ti plasmid and are, therefore, kanamycin-resistant (16). After 5 weeks, the roots were excised from the stem explants, cultured for three more weeks, and the fatty acid composition of total lipid extracts measured (Table 1). The *fad3* mutant transformed with only the vector had less 18:3 and more 18:2 fatty acid than did the wild type. Transformation of the *fad3* mutant with pTiDES3 produced roots that contained greater than wild-type amounts of 18:3 and lower amounts of 18:2. Transformation of the wild type with pTiDES3 resulted in even more 18:3. Thus, the *B. napus* cDNA in pTiDES3 functionally complements the *fad3* mutation in *Arabidopsis*, and the proportion of unsaturated lipids can be altered by affecting transcription of a putative desaturase gene.

Analysis of the nucleotide sequence of the cDNA insert in pBNDES3 revealed a 383-amino acid open reading frame that encodes a 44-kD polypeptide (Fig. 2). The NH<sub>2</sub>-terminal region does not exhibit the characteristics of a signal peptide (17) but the COOH-terminus contained the Lysyl residues three and five amino acids from the end that have been shown to be sufficient in animals for retention of membrane proteins in the endoplasmic reticulum (ER) membrane (18). Several strongly hydrophobic internal domains could be transmembrane domains. These characteristics suggest that the pBNDES3 cDNA encodes a membrane-bound protein located in the ER, which is in agreement with the available biochemical evidence concerning the localization of the  $\omega$ 3 desaturase encoded by the *fad3* locus (9). Comparison of the deduced amino acid sequence with the protein sequences contained in GenBank release 70 using the FASTA program (19)

indicated homology with the  $\omega$ 6 desaturase from the cyanobacterium *Synechocystis* (20), including a 12-residue sequence of which 10 residues were identical. No homology with other proteins could be detected. The sequence homology between the cyanobacterial  $\omega$ 6 desaturase and the *fad3* gene product raises the possibility that the  $\omega$ 3 and  $\omega$ 6 desaturases in higher plants may also have significant structural similarity.

*Note added in proof:* J. Browse, N. Yadav, and collaborators have cloned the *Arabidopsis fad3* gene by T-DNA tagging (21).

## REFERENCES AND NOTES

1. L. S. Leutwiler, B. Hough-Evans, E. M. Meyerowitz, *Mol. Gen. Genet.* 194, 15 (1984); R. E. Pruitt and E. M. Meyerowitz, *J. Mol. Biol.* 187, 169 (1986).
2. C. Chang, J. L. Bowman, A. W. DeJohn, E. S. Lander, E. M. Meyerowitz, *Proc. Natl. Acad. Sci. U.S.A.* 85, 6856 (1988).
3. H. G. Nam, et al., *Plant Cell* 1, 699 (1989).
4. E. Grill and C. R. Somerville, *Molec. Gen. Genet.* 226, 484 (1990).
5. E. R. Ward and G. C. Jen, *Plant Mol. Biol.* 14, 561 (1990).
6. P. L. Steponkus, D. V. Lynch, M. Uemura, *Philos. Trans. R. Soc. London Ser. B* 326, 571 (1990); P. Quinn, in *Plants and Temperature*, S. P. Long and F. I. Woodward, Eds. (Company of Biologists Ltd., Cambridge, 1988), pp. 237-259.
7. J. Harwood, *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 39, 101 (1988); J. Browse and C. R. Somerville, *ibid.* 42, 467 (1991).
8. C. Somerville and J. Browse, *Science* 252, 80 (1991).
9. B. Lemieux, M. Miquel, C. R. Somerville, J. Browse, *Theor. Appl. Genet.* 80, 234 (1990).
10. K. K. Niyogi and G. R. Fink, *Plant Cell* 4, 721 (1992).
11. J. Shanklin and C. R. Somerville, *Proc. Natl. Acad. Sci. U.S.A.* 88, 2510 (1991).
12. G. Chu, D. Vollrath, R. W. Davis, *Science* 234, 1582 (1986).
13. The six pairs of homologous genes from *Arabidopsis* and *B. napus* in GenBank release 70 had an average nucleotide sequence identity of 83%.
14. L. Moore, G. Warren, G. Strobel, *Plasmid* 2, 617 (1979); L. D. Owens and D. E. Cress, *Plant Physiol.* 77, 87 (1985).
15. Plants were grown for 4 weeks in continuous illumination at 23° to 25°C. Flowering stems, 10 to 20 cm in length, were surface-sterilized in 10% (v/v) Clorox, 0.02% Triton X-100, followed by four washes with sterile water. The apical 2 cm of inflorescences with flowers and unopened buds were blotted dry and the cut end dipped in an overnight L-broth culture of *A. tumefaciens*. After blotting to remove excess liquid, the stem section was pushed into the surface of agar-solidified medium so that the cut end was not in contact with the agar. The medium contained MS-salts (Gibco, Grand Island, NY), Gamborg's B5 vitamins (Gibco), 3% (w/v) sucrose, and 0.8% (w/v) agar. Following 1 day of incubation at 20° to 23°C in continuous light, the stem sections were transferred to the same medium containing cefotaxime (200 µg/ml). Hair roots were visible at the cut ends of the sections within 15 to 20 days after infection. Individual roots were excised, and cultured on agar or as liquid cultures in 24-well dishes, in the same medium plus kanamycin (50 µg/ml).
16. J. Martinez-Zapater, J. Schiefelbein, C. R. Somerville, unpublished results.
17. G. Von Heijne, *J. Mol. Biol.* 189, 239 (1986).
18. M. R. Jackson, T. Nilsson, P. A. Peterson, *EMBO J.* 9, 3153 (1990).
19. W. R. Pearson and D. J. Lipman, *Proc. Natl. Acad. Sci. U.S.A.* 85, 2444 (1988).
20. H. Wada, Z. Gombos, N. Murata, *Nature* 347, 200 (1990).
21. J. Browse et al., personal communication.

Fig. 2. Deduced amino acid sequence of the protein encoded by the *fad3* cDNA. The nucleotide sequence has been deposited in the GenBank database (L01418) and the clone is available from the *Arabidopsis* Biological Resource Center (22). Abbreviations for the amino acid residues are: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

22. *Arabidopsis* Biological Resource Center, 1735 Neil Avenue, Columbus, OH 43210.  
 23. J. Browne, N. Warwick, C. R. Somerville, C. R. Slack, *Biochem. J.* 235, 25 (1986).  
 24. We thank E. Meyerowitz for RFLP markers and mapping information, E. Ward and G. Jen for the YAC library, K. Niyogi and G. Fink for the ASA

marker, D. Murphy for the *B. napus* library, and E. Grill, B. Hauge, J. Schiefelbein, J. Martinez-Zapater, P. Gil, and K. Iba for advice. V.A., S.G., and B.L. were supported in part by fellowships from the European Molecular Biology Organization, National Institutes of Health, and the Natural Sciences and Engineering Research Council of

Canada, respectively. Supported by grants DCB8916311 from the National Science Foundation and DE-FG02-90ER20021 from the U.S. Department of Energy to C.R.S., and a grant from Hoechst AG to H.M.G.

29 June 1992; accepted 15 September 1992

## Dynamics of Ribozyme Binding of Substrate Revealed by Fluorescence-Detected Stopped-Flow Methods

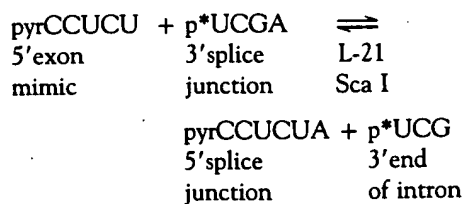
Philip C. Bevilacqua, Ryszard Kierzek, Kenneth A. Johnson, Douglas H. Turner\*

Fluorescence-detected stopped-flow and equilibrium methods have been used to study the mechanism for binding of pyrene (pyr)-labeled RNA oligomer substrates to the ribozyme (catalytic RNA) from *Tetrahymena thermophila*. The fluorescence of these substrates increases up to 25-fold on binding to the ribozyme. Stopped-flow experiments provide evidence that pyr experiences at least three different microenvironments during the binding process. A minimal mechanism is presented in which substrate initially base pairs to ribozyme and subsequently forms tertiary contacts in an RNA folding step. All four microscopic rate constants are measured for ribozyme binding of pyrCCUCU.

Recognition of the 5' exon for splicing of the ribosomal RNA precursor of *T. thermophila* involves base pairing of the exon sequence CUCUCU with part of an intron internal guide sequence (IGS), GGAGGG, to give a helix designated P1 (1). This process can be mimicked with oligonucleotides and truncated forms of the intron (2, 3). Tertiary interactions involving 2'-OH groups of substrate enhance this binding (4-6). Under conditions where all the ribozyme is active (6), we report transient kinetic studies indicating that tertiary contacts form after base pairing, and we provide the first rate constants for the dynamics of this RNA folding step (Fig. 1A).

Conjugation of pyrene to a 5' amino-modified ribose (Fig. 1B) provides a probe of rapid binding steps (7). Binding of pyrCUCU, pyrCCUCU, pyrCUCUCU, and pyrCCCUCU to the L-21 Sca I form of the ribozyme from *T. thermophila* (3) increases pyr fluorescence by factors of 25, 21, 8, and 4, respectively (8), consistent with expectations based on three-dimensional models of the binding site (9). All four pyr-labeled substrates reacted in single turnover nucleotidyl transfer reactions with <sup>32</sup>P-labeled (p\*) p\*UCGA, suggesting that their fluorescence enhancement results from binding

in the catalytic core (10). A typical reaction modeling the second step of splicing is



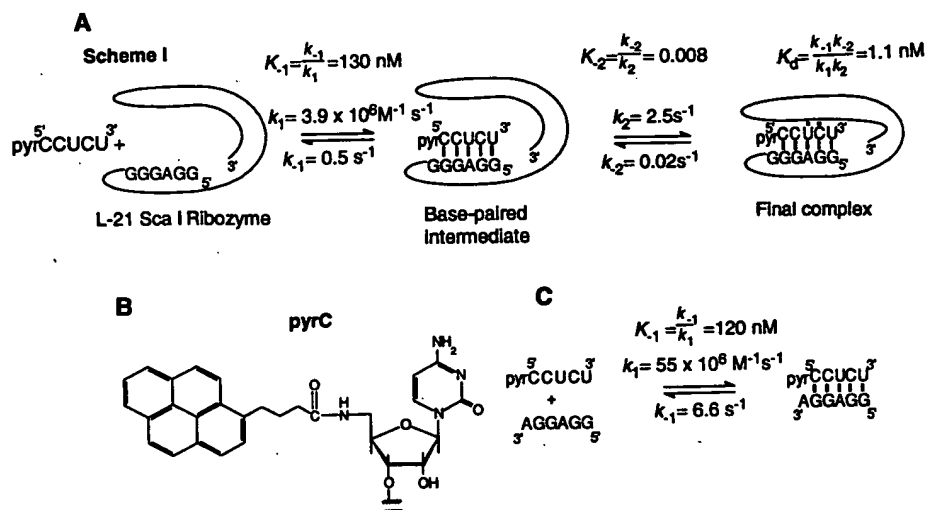
Rapid mixing, stopped-flow experiments

with L-21 Sca I and pyrCCUCU resulted in fast and slow rates for binding,  $1/\tau_1$  and  $1/\tau_2$ , respectively (Fig. 2). Traces of fluorescence versus time after mixing were fit to a single or double exponential as appropriate. Plots of rates versus [pyrCCUCU] give a straight line fit for the faster rate and a hyperbolic fit for the slower rate (Fig. 3A). This is consistent with two-step binding in which base pairing of pyrCCUCU and the IGS to form P1 occurs in the first step and uptake of P1 occurs in the second step (Fig. 1A, Scheme 1). The apparent enhancement of pyr fluorescence after tertiary folding is a unique observation for a nucleic acid (7). For Scheme 1 and substrate, S, in excess over L-21 Sca I (11)

$$\frac{1}{\tau_1} \approx k_1[S] + k_{-1} + k_2 + k_{-2} \quad (1)$$

$$\frac{1}{\tau_2} \approx \frac{k_1[S](k_2 + k_{-2}) + k_{-1}k_{-2}}{k_1[S] + k_{-1} + k_2 + k_{-2}} \quad (2)$$

where  $k_1$ ,  $k_{-1}$ ,  $k_2$ , and  $k_{-2}$  are the rate constants shown in Scheme 1. Initial estimates of rate constants were obtained by fitting data in Fig. 3A to Eqs. 1 and 2. Rate constants were optimized by computer simulation of traces (Table 1) (12). The value



**Fig. 1. (A)** Minimal mechanism (Scheme I) consistent with all the data for all four 5' exon mimics. The rate constants are for pyrCCUCU. The sketch of L-21 Sca I is not meant to give structural detail but to indicate that pyr is protected from solvent in both the intermediate and final state and less so as oligomer length increases. The sketch also depicts GGAGGG (IGS) as not completely accessible in the unbound L-21 Sca I. Lines indicate base pairing (1). Known tertiary hydrogen bonds involving 2' OH groups of substrate (5, 6) are indicated by bold dots. **(B)** Structure of the 5' end of pyr-modified oligomers (pyrC). **(C)** Minimal mechanism for pyrCUCU and pyrCCUCU binding to the IGS mimic GGAGGA. The rate constants are for pyrCCUCU.

P. C. Bevilacqua and D. H. Turner, Department of Chemistry, University of Rochester, Rochester, NY 14627.

R. Kierzek, Institute of Bio-organic Chemistry, Polish Academy of Sciences, 60-704 Poznan, Noskowskiego 12/14, Poland.

K. A. Johnson, Department of Molecular and Cell Biology, Pennsylvania State University, Althouse Laboratory, University Park, PA 16802.

\*To whom correspondence should be addressed.



Europäisches Patentamt  
European Patent Office  
Office européen des brevets

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

(11) Publication number: **0 420 358 A1**

(2)

## EUROPEAN PATENT APPLICATION

(21) Application number: **90202565.9**

(51) Int. Cl.<sup>4</sup>: **C12N 15/55, C12N 9/16,  
A23K 1/165, C12N 1/15,  
C12N 1/19**

(22) Date of filing: **27.09.90**

The applicant has filed a statement in accordance with Rule 28 (4) EPC (issue of a sample only to an expert). Accession number(s) of the deposit(s): CBS 155.89, CBS 513.88, CBS 213.88 and CBS 214.89.

(23) Priority: **27.09.89 EP 89202436  
17.08.90 EP 90202231**

(24) Date of publication of application:  
**03.04.91 Bulletin 91/14**

(26) Designated Contracting States:  
**AT BE CH DE DK ES FR GB GR IT LI LU NL SE**

(71) Applicant: **GIST-BROCADES N.V.  
Wateringseweg 1 P.O. Box 1  
NL-2600 MA Delft(NL)**

(72) Inventor: **van Gorcom, Robert Franciscus  
Maria**

**Liberlastraat 7  
NL-2622 DE Delft(NL)**

Inventor: **van Hartingsveldt, Willem**

**Laan van Altena 43  
NL-2613 DL Delft(NL)**

Inventor: **van Paridon, Peter Andreas  
Beekdaan 120**

**NL-2201 BG Noordwijk(NL)**

Inventor: **Veenstra, Annemarie Eveline  
Nierop 27**

**NL-2151 XH Nieuw Vennep(NL)**

Inventor: **Luiten, Rudolf Gijbertus Marie  
Akkerhoornbloem 26**

**NL-2317 KR Leiden(NL)**

Inventor: **Selten, Gerardus Cornelis Maria  
Sterrenweg 81**

**NL-2551 HZ Berkel en Rodenrijs(NL)**

(74) Representative: **Huygens, Arthur Victor, Dr. et  
al  
c/o Gist-Brocades N.V. Patent & Trademarks  
Department Wateringseweg 1 PO Box 1  
NL-2600 MA Delft(NL)**

(34) Cloning and expression of microbial phytase.

(37) A nucleotide sequence encoding phytase has been isolated and cloned. The coding sequence has been inserted into an expression construct which in turn has been inserted into a vector capable of transforming a microbial expression host. The transformed microbial hosts may be used to economically produce phytase on an industrial scale. The phytase produced via the present invention may be used in a variety of processes requiring the conversion of phytate to inositol and inorganic phosphate.



## CLONING AND EXPRESSION OF MICROBIAL PHYTASE

The present invention relates to the microbial production of phytase.

Background of the Invention

5

Phosphorus is an essential element for the growth of all organisms. In livestock production, feed must be supplemented with inorganic phosphorus in order to obtain a good growth performance of monogastric animals (e.g. pigs, poultry and fish).

10 In contrast, no inorganic phosphate needs to be added to the feedstuffs of ruminant animals. Microorganisms, present in the rumen, produce enzymes which catalyze the conversion of phytate (myo-inositolhexakis-phosphate) to inositol and inorganic phosphate.

Phytate occurs as a storage phosphorus source in virtually all feed substances originating from plants (for a review see: Phytic acid, chemistry and applications, E. Graf (ed.), Pilatus Press; Minneapolis, MN, U.S.A. (1986)). Phytate comprises 1-3% of all nuts, cereals, legumes, oil seeds, spores and pollen. Complex salts of phytic acid are termed phytin. Phytic acid is considered to be an anti-nutritional factor since it chelates minerals such as calcium, zinc, magnesium, iron and may also react with proteins, thereby decreasing the bioavailability of protein and nutritionally important minerals.

Phytate phosphorus passes through the gastro-intestinal tract of monogastric animals and is excreted in 20 the manure. Though some hydrolysis of phytate does occur in the colon, the thus-released inorganic phosphorus has no nutritional value since inorganic phosphorus is absorbed only in the small intestine. As a consequence, a significant amount of the nutritionally important phosphorus is not used by monogastric animals, despite its presence in the feed.

The excretion of phytate phosphorus in manure has further consequences. Intensive livestock production has increased enormously during the past decades. Consequently, the amount of manure produced has increased correspondingly and has caused environmental problems in various parts of the world. This is due, in part, to the accumulation of phosphate from manure in surface waters which has caused eutrophication.

The enzymes produced by microorganisms, that catalyze the conversion of phytate to inositol and 30 inorganic phosphorus are broadly known as phytases. Phytase producing microorganisms comprise bacteria such as Bacillus subtilis (V.K. Paver and V.J. Jagannathan (1982) J. Bacteriol. 151, 1102-1108) and Pseudomonas (D.J. Cosgrove (1970) Austral. J. Biol. Sci. 23, 1207-1220); yeasts such as Saccharomyces cerevisiae (N.R. Nayini and P. Markakis (1984) Lebensmittel Wissenschaft und Technologie 17, 24-26); and fungi such as Aspergillus terreus (K. Yamada, Y. Minoda and S. Yamamoto (1986) Agric. Biol. Chem. 32, 1275-1282). Various other Aspergillus species are known to produce phytase, of which, the phytase produced by Aspergillus ficuum has been determined to possess one of the highest levels of specific activity, as well as having better thermostability than phytases produced by other microorganisms (unpublished observations).

The concept of adding microbial phytase to the feedstuffs of monogastric animals has been previously 40 described (Ware, J.H., Bluff, L. and Shieh, T.R. (1967) U.S. Patent No. 3,297,548; Nelson, T.S., Shieh, T.R., Wodzinski, R.J. and Ware, J.H. (1971) J. Nutrition 101, 1289-1294). To date, however, application of this concept has not been commercially feasible, due to the high cost of the production of the microbial enzymes (Y.W. Han (1989) Animal Feed Sci. & Technol. 24, 345-350). For economic reasons, inorganic phosphorus is still added to monogastric animal feedstuffs.

45 Microbial phytases have found other industrial uses as well. Exemplary of such utilities is an industrial process for the production of starch from cereals such as corn and wheat. Waste products comprising e.g. corn gluten feeds from such a wet milling process are sold as animal feed. During the steeping process phytase may be supplemented. Conditions ( $T \approx 50^{\circ}\text{C}$  and  $\text{pH} = 5.5$ ) are ideal for fungal phytases (see e.g. European Patent Application 0 321 004 to Alko Ltd.). Advantageously, animal feeds derived from the waste products of this process will contain phosphate instead of phytate.

50 It has also been conceived that phytases may be used in soy processing (see Finase™ Enzymes By Alko, a product information brochure published by Alko Ltd., Rajamäki, Finland). Soybean meal contains high levels of the anti-nutritional factor phytate which renders this protein source unsuitable for application in baby food and feed for fish, calves and other non-ruminants. Enzymatic upgrading of this valuable protein source improves the nutritional and commercial value of this material.

Other researchers have become interested in better characterizing various phytases and improving procedures for the production and use of these phytases. Ullah has published a procedure for the purification of phytase from wild-type *Aspergillus ficuum*, as well as having determined several biochemical parameters of the product obtained by this purification procedure (Ullah, A. (1988a) *Preparative Biochem.* 18, 443-458). Pertinent data obtained by Ullah is presented in Table 1, below.

The amino acid sequence of the N-terminus of the *A. ficuum* phytase protein has twice been disclosed by Ullah: Ullah, A. (1987) *Enzyme and Engineering conference IX*, October 4-8, 1987, Santa Barbara, California (poster presentation); and Ullah, A. (1988b) *Prep. Biochem.* 18, 459-471. The amino acid sequence data obtained by Ullah is reproduced in Figure 1A, sequence E, below.

Several interesting observations may be made from the disclosures of Ullah. First of all, the "purified" preparation described in Ullah (1988a and 1988b) consists of two protein bands on SDS-PAGE. We have found, however, that phytase purified from *A. ficuum* contains a contaminant and that one of the bands found on SDS-PAGE, identified by Ullah as a phytase, is originating from this contaminant.

This difference is also apparent from the amino acid sequencing data published by Ullah (1987, 1988b; compare Figure 1A, sequences A and B with sequence C). We have determined, in fact, that one of the amino acid sequences of internal peptides of phytase described by Ullah (see Figure 1B, sequence E) actually belongs to the contaminating 100 kDa protein (Figure 1C) which is present in the preparation obtained via the procedure as described by Ullah, and seen as one of the two bands on SDS-PAGE (Ullah, 1988a and 1988b). Ullah does not recognize the presence of such a contaminating protein, and instead identifies it as another form of phytase. The presence of such contamination, in turn, increases the difficulty in selecting and isolating the actual nucleotide sequence encoding phytase activity. Furthermore, the presence of the contamination lowers the specific activity value of the protein tested.

Further regarding the sequence published by Ullah, it should be noted that the amino acid residue at position 12, has been disclosed by Ullah to be glycine. We have consistently found using protein and DNA sequencing techniques, that this residue is not a glycine but is in fact a cysteine (see Figures 6 and 8).

Finally, Ullah discloses that phytase is an 85 kDa protein, with a molecular weight after deglycosylation of 61.7 kDa (Ullah, 1988b). This number, which is much lower than the earlier reported 76 kDa protein (Ullah, A. and Gibson, D. (1988) *Prep. Biochem.* 17(1), 63-91) was based on the relative amount of carbohydrates released by hydrolysis, and the apparent molecular weight of the native protein on SDS-PAGE. We have found, however, that glycosylated phytase has a single apparent molecular weight of 85 kDa, while the deglycosylated protein has an apparent molecular weight in the range of 48 - 56.5 kDa, depending on the degree of deglycosylation.

Mullaney et al. (Filamentous Fungi Conference, April, 1987, Pacific Grove, California (poster presentation) also disclose the characterization of phytase from *A. ficuum*. However, this report also contains mention of two protein bands on SDS-PAGE, one of 85 kDa, and one of 100 kDa, which were present in the "purified" protein preparation. These protein bands are both identified by the authors as being forms of phytase. A method for transforming microbial hosts is proposed, but has not been reported. The cloning and isolation of the DNA sequence encoding phytase has not been described.

It will be appreciated that an economical procedure for the production of phytase will be of significant benefit to, inter alia, the animal feed industry. One method of producing a more economical phytase would be to use recombinant DNA techniques to raise expression levels of the enzyme in various microorganisms known to produce high levels of expressed peptides or proteins. To date, however, the isolation and cloning of the DNA sequence encoding phytase activity has not been published.

#### Summary of the Invention

The present invention provides a purified and isolated DNA sequence coding for phytase. The isolation and cloning of this phytase encoding DNA sequence has been achieved via the use of specific oligonucleotide probes which were developed especially for the present invention. Preferred DNA sequences encoding phytases are obtainable from fungal sources, especially filamentous fungi of the genus *Aspergillus*.

It is another object of the present invention to provide a vector containing an expression construct which further contains at least one copy of at least one, preferably homologous DNA sequence encoding phytase, operably linked to an appropriate regulatory region capable of directing the high level expression of peptides or proteins having phytase activity in a suitable expression host.

The expression construct provided by the present invention may be inserted into a vector, preferably a

plasmid, which is capable of transforming a microbial host cell and integrating into the genome.

It is a further object of the present invention to provide a transformant, preferably, a microbial host which has been transformed by a vector as described in the preceding paragraph. The transformed hosts provided by the present invention are filamentous fungi of the genera Aspergillus, Trichoderma, Mucor and Penicillium, yeasts of the genera Kluyveromyces and Saccharomyces or bacteria of the genus Bacillus. Especially preferred expression hosts are filamentous fungi of the genus Aspergillus. The transformed hosts are capable of producing high levels of recombinant phytase on an economical, industrial scale.

In other aspects, the invention is directed to recombinant peptides and proteins having phytase activity in glycosylated or unglycosylated form; to a method for the production of said unglycosylated peptides and proteins; to peptides and proteins having phytase activity which are free of impurities; and to monoclonal antibodies reactive with these recombinant or purified proteins.

A comparison of the biochemical parameters of the purified wild-type A. ficuum phytase as obtained by Ullah, against the further purified wild-type A. ficuum phytase, obtained via the present invention, is found in Table 1, below. Of particular note is the specific activity data wherein it is shown that the purified protein which we have obtained has twice the specific activity of that which was published by Ullah.

The present invention further provides nucleotide sequences encoding proteins exhibiting phytase activity, as well as amino acid sequences of these proteins. The sequences provided may be used to design oligonucleotide probes which may in turn be used in hybridization screening studies for the identification of phytase genes from other species, especially microbial species, which may be subsequently isolated and cloned.

The sequences provided by the present invention may also be used as starting materials for the construction of "second generation" phytases. "Second generation" phytases are phytases, altered by mutagenesis techniques (e.g. site-directed mutagenesis), which have properties that differ from those of wild-type phytases or recombinant phytases such as those produced by the present invention. For example, the temperature or pH optimum, specific activity or substrate affinity may be altered so as to be better suited for application in a defined process.

Within the context of the present invention, the term phytase embraces a family of enzymes which catalyze reactions involving the removal of inorganic phosphorous from various myoinositol phosphates.

Phytase activity may be measured via a number of assays, the choice of which is not critical to the present invention. For purposes of illustration, phytase activity may be determined by measuring the amount of enzyme which liberates inorganic phosphorous from 1.5 mM sodium phytate at the rate of 1  $\mu\text{mol/min}$  at 37°C and at pH 5.50.

It should be noted that the term "phytase" as recited throughout the text of this specification is intended to encompass all peptides and proteins having phytase activity. This point is illustrated in Figure 1A which compares sequences A and B (sequences which have been obtained during the course of the present work) with sequence C (published by Ullah, 1988b). The Figure demonstrates that proteins may be obtained via the present invention which lack the first four amino acids (the protein of sequence A lacks the first seven amino acids) of the mature A. ficuum phytase protein. These proteins, however, retain phytase activity. The complete amino acid sequence of the phytase protein, as deduced from the corresponding nucleotide sequence, is shown in Figure 8.

Phytases produced via the present invention may be applied to a variety of processes which require the conversion of phytate to inositol and inorganic phosphate.

For example, the production of phytases according to the present invention will reduce production costs of microbial phytases in order to allow its economical application in animal feed which eventually will lead to an *in vivo* price/performance ratio competitive with inorganic phosphate. As a further benefit, the phosphorus content of manure will be considerably decreased.

It will be appreciated that the application of phytases, available at a price competitive with inorganic phosphate, will increase the degrees of freedom for the compound feed industry to produce a high quality feed. For example, when feed is supplemented with phytase, the addition of inorganic phosphate may be omitted and the contents of various materials containing phytate may be increased.

In addition to use in animal feeds and soy processing as discussed above, the phytase obtained via the present invention may also be used in diverse industrial applications such as:

- liquid feed for pigs and poultry. It has become common practice to soak feed for several hours prior to feeding. During this period the enzyme will be able to convert phytate to inositol and inorganic phosphate;
- an industrial process for the production of inositol or inositol-phosphates from phytate;
- other industrial processes using substrates that contain phytate such as the starch industry and in fermentation industries, such as the brewing industry. Chelation of metal ions by phytate may cause these minerals to be unavailable for the production microorganisms. Enzymatic hydrolysis of phytate prevents

these problems.

These and other objects and advantages of the present invention will become apparent from the following detailed description.

5

### Brief description of the Figures

#### Figure 1.

10 A. N-terminal amino acid sequences as determined for purified phytase. The amino acid sequences labeled A and B are provided by the present invention, and originate from the phytase subforms with isoelectric points of 5.2 and 5.4, respectively. Sequence C is cited from Ullah (1987, 1988b, supra). The amino acid residue located at position 12 of sequences A and B has been determined by the present invention not to be a glycine residue. [\* denotes no unambiguous identification. " denotes no residue detected.]

15 B. N-terminal amino acid sequences of CNBr-cleaved internal phytase fragments. The amino acid sequences labeled A and B (apparent molecular weight approximately 2.5 kDa and 36 kDa peptides, respectively) are provided by the present invention. Sequences C through E are cited from Ullah (1988b, supra).

20 C. N-terminal amino acid sequence of a 100 kDa protein which has been found by the present invention to be present in crude phytase samples.

#### Figure 2.

A. Oligonucleotide probes designed on basis of the data from Figure 1A, peptides A through E.

B. Oligonucleotide probes designed on the basis of the data from Figure 1B, peptides A and B.

25 Figure 3. Oligonucleotide probes used for the isolation of the gene encoding the acid-phosphatase.

Figure 4. Restriction map of bacteriophage lambda AF201 containing the phytase locus of *A. ficuum*. The arrow indicates the position of the phytase gene and the direction of transcription. Clone # shows the subclones derived with indicated restriction enzymes from phage AF201 in pAN 8-1 (for pAF 28-1) and in pUC19 (for all other subclones).

30 Figure 5. Physical map of pAF 1-1. The 10 kb Bam HI fragment, inserted in pUC19, contains the entire gene encoding acid phosphatase from *A. ficuum*.

Figure 6. Compilation of the nucleotide sequences of plasmids pAF 2-3, pAF 2-6, and pAF 2-7 encompassing the chromosomal phytase gene locus. The phytase coding region is located from nucleotide position 210 to position 1713; an intron is present in the chromosomal gene from nucleotide position 254 to position 355. Relevant features such as restriction sites, the phytase start and stop codons, and the Intron position are indicated.

Figure 7. Detailed physical map of the sequenced phytase chromosomal locus; the arrows indicate the location of the two exons of the phytase coding region.

Figure 8. Nucleotide sequence of the translated region of the phytase cDNA fragment and the derived amino acid sequence of the phytase protein; the start of the mature phytase protein is indicated as position +1. The amino-terminus of the 36 kDa internal protein fragment is located at amino acid position 241, whereas the 2.5 kDa protein fragment starts at amino acid position 390.

Figure 9. Physical map of the phytase expression cassette pAF 2-2S. Arrows indicate the direction of transcription of the genes.

45 Figure 10. IEF-PAGE evidence of the overexpression of phytase in an *A. ficuum* NRRL 3135 transformant. Equal volumes of culture supernatant of *A. ficuum* (lane 1) and transformant pAF 2-2S SP7 (lane 2), grown under identical conditions, were analysed on a Phast-System (Pharmacia) IEF-PAGE gel in the pH-range of 4.5-6. For comparison, a sample of *A. ficuum* phytase, purified to homogeneity was included either separately (lane 4), or mixed with a culture supernatant (lane 3). The gels were either stained with a phosphatase stain described in the text (A), or with a general protein stain (Coomassie Brilliant Blue, B). The phytase bands are indicated by an asterisk.

Figure 11. IEF-PAGE evidence for the overexpression of phytase in *A. niger* CBS 513.88 transformants. Equal volumes of culture supernatants of the *A. niger* parent strain (lane 1), or the transformants pAF 2-2S #8 (lane 2), pFYT3 #205 (lane 3) & #282 (lane 4) were analysed by IEF-PAGE as described in the legend of Fig. 10. The gels were either stained by a general phosphatase activity stain (A) or by a general protein stain (B). Phytase bands are indicated by an asterisk.

55 Figure 12. Physical map of pAB 8-1. The 14.5 kb Hin dIII DNA insert in pUC19 contains the entire glucoamylase (AG) locus from *A. niger*.

Figure 13. A schematic view of the generation of AG promoter/phytase gene fusions by the polymerase chain reaction (PCR). The sequences of all oligonucleotide primers used are indicated in the text.

Figure 14. Physical map of the phytase expression cassette pAF 2-2SH.

Figure 15. Physical maps of the intermediate constructs pXXFYT1, pXXFYT2 and the phytase expression cassettes pXXFYT3, wherein XX indicates the leader sequence (L). In p18FYT# and p24FYT#, respectively the 18 aa and the 24 aa AG leader sequence are inserted whereas in pFYT#, the phytase leader is used.

Figure 16. Physical map of plasmid pFYT3ΔamdS.

Figure 17. Physical map of plasmid pFYT3INT.

Figure 18. Physical map of the phytase/AG replacement vector pREPFYT3.

Figure 19. Autoradiographs of chromosomal DNA, digested with Pvu II (A) and Bam HI (B) and hybridized with the <sup>32</sup>P-labeled *A. ficuum* phytase cDNA as probe of the microbial species *S. cerevisiae* (lane 2); *B. subtilis* (lane 3); *K. lactis* (lane 4); *P. crysogenum* (lane 5); *P. aeruginosa* (lane 6); *S. lividans* (lane 7); *A. niger* 1 μg (lane 8); *A. niger* 5 μg (lane 9); blank (lane 10); *C. thermocellum* (lane 11). Lane 1: marker DNA.

#### Detailed Description of the Invention

The cloning of the genes encoding selected proteins produced by a microorganism can be achieved in various ways. One method is by purification of the protein of interest, subsequent determination of its N-terminal amino acid sequence and screening of a genomic library of said micro-organism using a DNA oligonucleotide probe based on said N-terminal amino acid sequence. Examples of the successful application of this procedure are the cloning of the Isopenicillin N-synthetase gene from *Cephalosporium acremonium* (S.M. Samson et al. (1985) *Nature* 318, 191-194) and the isolation of the gene encoding the TAKA amylase for *Aspergillus oryzae* (Boel et al. (1986) EP-A-0238023).

Using this procedure, an attempt has been made to isolate the *Aspergillus ficuum* gene encoding phytase. The protein has been purified extensively, and several biochemical parameters have been determined. The data obtained have been compared to the data published by Ullah (1988a). Both sets of data are given in Table 1, below.

Table 1

Biochemical parameters of purified wild-type A. ficuum  
phytase

<u>Parameter</u>	<u>Present invention</u>	<u>Ullah</u>
Specific activity*	100 U/mg protein	50 U/mg protein
Purity: SDS-PAGE	85 kDa	85 / 100 kDa
: IEF-PAGE	3 or 4 bands	not done
K <sub>m</sub> (Affinity constant)	250 $\mu$ M	40 $\mu$ M
Specificity for:		
Inositol-1-P	not active	not active
Inositol-2-P	K <sub>m</sub> = 3.3mM	5% activity
pH optimum	2.5 and 5.5	2.5 and 5.5
Temp. optimum (°C)	50	58
MW (kDa)**	85	85 and 100
MW (unglycosylated)**	56.5	61.7
Isoelectric Point***	5.0-5.4	4.5

\* Phytase activity is measured by Ullah at 58°C rather than at 37°C. A unit of phytase activity is defined as that amount of enzyme which liberates inorganic phosphorus from 1.5 mM sodium phytate at the rate of 1  $\mu$ mol/min at 37°C and at pH 5.50. To compare the fermentation yields and the specific activities, the activities disclosed by Ullah were corrected for the temperature difference. The correction is based on the difference in phytase activity measured at 37°C and at 58°C as shown in Table III of Ullah (1988b).

\*\* Apparent Molecular Weight as determined by SDS-PAGE.

\*\*\* As determined by IEF-PAGE

In order to isolate the gene encoding phytase, a first set of oligonucleotide probes was designed according to the above-described method (Fig. 2A). The design of these probes was based on the amino acid sequence data. As a control for the entire procedure, similar steps were taken to isolate the gene encoding acid-phosphatase, thereby using the protein data published by Ullah and Cummins ((1987) Prep. Biochem. 17, 397-422). For acid-phosphatase, the corresponding gene has been isolated without difficulties. However, for phytase, the situation appeared to be different. Despite many attempts in which probes derived from the N-terminal amino acid sequence were used, no genomic DNA fragments or clones from the genomic library could be isolated which could be positively identified to encompass the gene encoding phytase.

To overcome this problem, the purified phytase was subjected to CNBr-directed cleavage and the

resulting protein fragments were isolated. The N-terminal amino acid sequences of these fragments were determined (Fig. 1B), and new oligonucleotide probes were designed, based on the new data (Fig. 2B). Surprisingly, the new oligonucleotide probes did identify specific DNA fragments and were suited to unambiguously identify clones from a genomic library. No cross hybridization was observed between the new clones or DNA fragments isolated there from, and the first set of oligonucleotide probes or the clones isolated using the first set of probes.

It will be appreciated that this second set of probes may also be used to identify the coding sequences of related phytases.

The newly isolated clones were used as probes in Northern blot hybridizations. A discrete mRNA could only be detected when the mRNA was isolated from phytase producing mycelium. When RNA from non-phytase producing mycelium was attempted, no hybridization signal was found. The mRNA has a size of about 1800 b, theoretically yielding a protein having a maximal molecular weight of about 60 kDa. This value corresponds to the molecular weight which has been determined for the non-glycosylated protein, and the molecular weight of the protein as deduced from the DNA sequence.

Moreover, when introduced into a fungal cell by transformation, an increase in phytase activity could be demonstrated. This indicates conclusively that the nucleotide sequence encoding phytase has indeed been isolated. The amino acid sequences which have been determined for the purified phytase enzyme, and for the CNBr fragments obtained therefrom, concur with the amino acid sequence deduced from the sequence which was determined for the cloned gene. The nucleotide sequence and the deduced amino acid sequence are given in Figures 6 and 8, and further illustrate the cloned sequence encoding phytase.

The isolation of the nucleotide sequence encoding phytase enables the economical production of phytase on an industrial scale, via the application of modern recombinant DNA techniques such as gene amplification, the exchange of regulatory elements such as e.g. promoters, secretional signals, or combinations thereof.

Accordingly, the present invention also comprises a transformed expression host capable of the efficient expression of high levels of peptides or proteins having phytase activity and, if desired, the efficient expression of acid phosphatases as well. Expression hosts of interest are filamentous fungi selected from the genera Aspergillus, Trichoderma, Mucor and Penicillium, yeasts selected from the genera Kluyveromyces and Saccharomyces and bacteria of the genus Bacillus. Preferably, an expression host is selected which is capable of the efficient secretion of their endogenous proteins.

Of particular interest are industrial strains of Aspergillus, especially niger, ficuum, awamori or oryzae. Alternatively, Trichoderma reesei, Mucor miehei, Kluyveromyces lactis, Saccharomyces cerevisiae, Bacillus subtilis or Bacillus licheniformis may be used.

The expression construct will comprise the nucleotide sequences encoding the desired enzyme product to be expressed, usually having a signal sequence which is functional in the host and provides for secretion of the product peptide or protein.

Various signal sequences may be used according to the present invention. A signal sequence which is homologous to the cloned nucleotide sequence to be expressed may be used. Alternatively, a signal sequence which is homologous or substantially homologous with the signal sequence of a gene at the target locus of the host may be used to facilitate homologous recombination. Furthermore, signal sequences which have been designed to provide for improved secretion from the selected expression host may also be used. For example, see Von Heyne (1983) Eur. J. Biochem. 133, 17-21; and Perlman and Halverson (1983) J. Mol. Biol. 167, 391-409. The DNA sequence encoding the signal sequence may be joined directly through the sequence encoding the processing signal (cleavage recognition site) to the sequence encoding the desired protein, or through a short bridge, usually fewer than ten codons.

Preferred secretional signal sequences to be used within the scope of the present invention are the signal sequence homologous to the cloned nucleotide sequence to be expressed, the 18 amino acid glucoamylase (AG) signal sequence and the 24 amino acid glucoamylase (AG) signal sequence, the latter two being either homologous or heterologous to the nucleotide sequence to be expressed.

The expression product, or nucleotide sequence of interest may be DNA which is homologous or heterologous to the expression host.

"Homologous" DNA is herein defined as DNA originating from the same genus. For example, Aspergillus is transformed with DNA from Aspergillus. In this way it is possible to improve already existing properties of the fungal genus without introducing new properties, which were not present in the genus before.

"Heterologous" DNA is defined as DNA originating from more than one genus, i.e., as follows from the example given in the preceding paragraph, DNA originating from a genus other than Aspergillus, which is then expressed in Aspergillus.

Nucleotide sequences encoding phytase activity are preferably obtained from a fungal source. More preferred are phytase encoding nucleotide sequences obtained from the genus Aspergillus. Most preferred sequences are obtained from the species Aspergillus ficuum or Aspergillus niger.

The region 5' to the open reading frame in the nucleotide sequence of interest will comprise the transcriptional initiation regulatory region (or promoter). Any region functional in the host may be employed, including the promoter which is homologous to the phytase-encoding nucleotide sequence to be expressed. However, for the most part, the region which is employed will be homologous with the region of the target locus. This has the effect of substituting the expression product of the target locus with the expression product of interest. To the extent that the level of expression and secretion of the target locus encoded protein provides for efficient production, this transcription initiation regulatory region will normally be found to be satisfactory. However, in some instances, one may wish a higher level of transcription than the target locus gene or one may wish to have inducible expression employing a particular inducing agent. In those instances, a transcriptional initiation regulatory region will be employed which is different from the region in the target locus gene. A large number of transcriptional initiation regulatory regions are known which are functional in filamentous fungi. These regions include those from genes encoding glucoamylase (AG), fungal amylase, acid phosphatase, GAPDH, Trp C, Amd S, Alc A, Ald A, histone H2A, Pyr 4, Pyr G, isopenicillin N synthetase, PGK, acid protease, acyl transferase, and the like.

The target locus will preferably encode a highly expressed protein gene, i.e., a gene whose expression product is expressed to a concentration of at least about 0.1 g/l at the end of the fermentation process. The duration of this process may vary inter alia on the protein product desired. As an example of such a gene, the gene encoding glucoamylase (AG) is illustrative. Other genes of interest include fungal  $\alpha$ -amylase, acid phosphatase, protease, acid protease, lipase, phytase and cellobiohydrolase. Especially preferred target loci are the glucoamylase gene of A. niger, the fungal amylase gene of A. oryzae, the cellobiohydrolase genes of T. reesei, the acid protease gene of Mucor miehei, the lactase gene of Kluyveromyces lactis or the invertase gene of Saccharomyces cerevisiae.

The transcriptional termination regulatory region may be from the gene of interest, the target locus, or any other convenient sequence. Where the construct includes further sequences of interest downstream (in the direction of transcription) from the gene of interest, the transcriptional termination regulatory region, if homologous with the target locus, should be substantially smaller than the homologous flanking region.

A selection marker is usually employed, which may be part of the expression construct or separate from the expression construct, so that it may integrate at a site different from the gene of interest. Since the recombinant molecules of the invention are preferably transformed to a host strain that can be used for industrial production, selection markers to monitor the transformation are preferably dominant selection markers, i.e., no mutations have to be introduced into the host strain to be able to use these selection markers. Examples of these are markers that enable transformants to grow on defined nutrient sources (e.g. the A. nidulans amd S gene enables A. niger transformants to grow on acetamide as the sole nitrogen source) or markers that confer resistance to antibiotics (e. g., the ble gene confers resistance to phleomycin or the hph gene confers resistance to hygromycin B).

The selection gene will have its own transcriptional and translational initiation and termination regulatory regions to allow for independent expression of the marker. A large number of transcriptional initiation regulatory regions are known as described previously and may be used in conjunction with the marker gene. Where antibiotic resistance is employed, the concentration of the antibiotic for selection will vary depending upon the antibiotic, generally ranging from about 30 to 300  $\mu$ g/ml of the antibiotic.

The various sequences may be joined in accordance with known techniques, such as restriction, joining complementary restriction sites and ligating, blunt ending by filling in overhangs and blunt ligation, Bal 31 resection, primer repair, in vitro mutagenesis, or the like. Polylinkers and adapters may be employed, when appropriate, and introduced or removed by known techniques to allow for ease of assembly of the expression construct. At each stage of the synthesis of the construct, the fragment may be cloned, analyzed by restriction enzyme, sequencing or hybridization, or the like. A large number of vectors are available for cloning and the particular choice is not critical to this invention. Normally, cloning will occur in E. coli.

The flanking regions may include at least part of the open reading frame of the target locus, particularly the signal sequence, the regulatory regions 5' and 3' of the gene of the target locus, or may extend beyond the regulatory regions. Normally, a flanking region will be at least 100 bp, usually at least 200 bp, and may be 500 bp or more. The flanking regions are selected, so as to disrupt the target gene and prevent its expression. This can be achieved by inserting the expression cassette (comprising the nucleotide sequence to be expressed and optionally including additional elements such as a signal sequence, a transcriptional initiation regulatory region sequence and/or a transcriptional termination regulatory region sequence) into



the open reading frame proximal to the 5' region, by substituting all or a portion of the target gene with the expression construct, or by having the expression construct intervene between the transcriptional initiation regulatory region at the target locus and the open reading frame. As already indicated, where the termination regulatory region is homologous with the region at the target locus, the 3'-flanking region should be substantially larger than a termination regulatory region present in the construct.

The present invention also provides the starting material for the construction of 'second-generation' phytases, i.e. phytase enzymes with properties that differ from those of the enzyme isolated herein. Second-generation phytases may have a changed temperature or pH optimum, a changed specific activity or affinity for its substrates, or any other changed quality that makes the enzyme more suited for application in a defined process. *E. coli* is the best host for such mutagenesis (e. g. site-directed mutagenesis). Since *E. coli* lacks the splicing machinery for the removal of introns which might be present in the phytase gene, a cDNA clone of phytase is the sequence of choice to be expressed in *E. coli*. This cDNA sequence can be readily mutated by procedures well known in the art, after which the mutated gene may be introduced into the desired expression constructs.

The construct may be transformed into the host as the cloning vector, either linear or circular, or may be removed from the cloning vector as desired. The cloning vector is preferably a plasmid. The plasmid will usually be linearized within about 1 kbp of the gene of interest. Preferably, the expression construct for the production of the phytases of the present invention will be integrated into the genome of the selected expression host.

A variety of techniques exist for transformation of filamentous fungi. These techniques include protoplast fusion or transformation, electroporation and micro-projectile firing into cells. Protoplast transformation has been found to be successful and may be used with advantage.

Mycelium of the fungal strain of interest is first converted to protoplasts by enzymatic digestion of the cell wall in the presence of an osmotic stabilizer such as KCl or sorbitol. DNA uptake by the protoplasts is aided by the addition of  $\text{CaCl}_2$  and a concentrated solution of polyethylene glycol, the latter substance causing aggregation of the protoplasts, by which process the transforming DNA is included in the aggregates and taken up by the protoplasts. Protoplasts are subsequently allowed to regenerate on solid medium, containing an osmotic stabilizer and, when appropriate, a selective agent, for which the resistance is encoded by the transforming DNA.

After selecting for transformants, the presence of the gene of interest may be determined in a variety of ways. By employing antibodies, where the expression product is heterologous to the host, one can detect the presence of expression of the gene of interest. Alternatively, one may use Southern or Northern blots to detect the presence of the integrated gene or its transcription product.

Amplification of the nucleotide sequence or expression construct of interest may be achieved via standard techniques such as, the introduction of multiple copies of the construct in the transforming vector or the use of the *amdS* gene as a selective marker (e.g. Weinans et al. (1985) *Current Genetics*, 9, 361-368). The DNA sequence to be amplified may comprise DNA which is either homologous or heterologous to the expression host, as discussed above.

The cells may then be grown in a convenient nutrient medium. Low concentrations of a protease inhibitor may be employed, such as phenylmethylsulfonyl fluoride,  $\alpha$ 2-macroglobulins, pepstatin, or the like. Usually, the concentration will be in the range of about 1  $\mu\text{g/ml}$  to 1  $\text{mg/ml}$ . The protease gene(s) may be inactivated in order to avoid or reduce degradation of the desired protein.

The transformants may be grown in either batch or continuous fermentation reactors, where the nutrient medium is isolated and the desired product extracted.

Various methods for purifying the product, if necessary, may be employed, such as chromatography (e. g., HPLC), solvent-solvent extraction, electrophoresis, combinations thereof, or the like.

The present invention also provides a downstream processing method in which the fermentation broth (optionally purified) is filtered, followed by a second germ-free filtration, after which the filtered solution is concentrated. The thus-obtained liquid concentrate may be used as follows:

a) Phytase and other proteins may be precipitated from the liquid concentrate by adding acetone to a final volume of 60% (v/v) under continuous stirring. The precipitate may be dried in a vacuum at 35° C. After grinding the dry powder, the enzyme product may be used as such for application experiments. Recovery yields are about 90%.

b) The liquid concentrate may be spray-dried using conventional spray-drying techniques. Recovery yields vary from 80 to 99%.

c) The liquid concentrate may be mixed with carrier materials such as wheat bran. The thus-obtained mixture may be dried in a spray tower or in a fluid bed.

d) The liquid concentrate may be osmotically stabilized by the addition of e.g. sorbitol. A preservative

such as benzoic acid may be added to prevent microbial contamination.

All four formulations may be sold to premix manufacturers, compound feed industries, other distributors and farmers.

The examples herein are given by way of illustration and are in no way intended to limit the scope of the present invention. It will be obvious to those skilled in the art that the phytase gene of the invention can be used in heterologous hybridization experiments, directed to the isolation of phytase encoding genes from other micro-organisms.

#### Example 1

##### Fermentation of *A. ficuum* NRRL 3135

*Aspergillus ficuum* strain NRRL 3135 was obtained from the Northern Region Research Lab, USDA, 1815 North University Street, Peoria, Illinois, USA. Fungal spore preparations were made following standard techniques.

Spores and subsequently cells were transferred through a series of batch fermentations in Erlenmeyer flasks to a 10 l fermentor. After growth in batch culture, the contents of this fermentor were used as inoculum for a final 500 liter batch fermentation.

The media used contains: 91 g/l corn starch (BDH Chemicals Ltd.); 38 g/l glucose.H<sub>2</sub>O; 0.6 g/l MgSO<sub>4</sub>.7H<sub>2</sub>O; 0.6 g/l KCl; 0.2 g/l FeSO<sub>4</sub>.7H<sub>2</sub>O and 12 g/l KNO<sub>3</sub>. The pH was maintained at 4.6 ± 0.3 by automatic titration with either 4N NaOH or 4N H<sub>2</sub>SO<sub>4</sub>.

Cells were grown at 28 °C at an automatically controlled dissolved oxygen concentration of 25% air saturation. Phytase production reached a maximum level of 5-10 U/ml after 10 days of fermentation.

#### Example 2

##### Purification and characterization of *A. ficuum* phytase

###### A. Phytase activity assay

100 µl of broth filtrate (diluted when necessary) or supernatant or 100 µl of demiwater as reference are added to an incubation mixture having the following composition:

- 0.25 M sodium acetate buffer pH 5.5, or
- glycine HCL-buffer pH 2.5
- 1mM phytic acid, sodium salt
- demiwater up to 900 µl

The resulting mixture is incubated for 30 minutes at 37 °C. The reaction is stopped by the addition of 1 ml of 10% TCA (trichloroacetic acid). After the reaction has terminated, 2 ml of reagent (3.66 g of FeSO<sub>4</sub>.7H<sub>2</sub>O in 50 ml of ammonium molybdate solution (2.5 g (NH<sub>4</sub>)<sub>6</sub>Mo<sub>7</sub>O<sub>24</sub>.4H<sub>2</sub>O and 8 ml H<sub>2</sub>SO<sub>4</sub>, diluted up to 250 ml with demiwater)) is added.

The intensity of the blue color is measured spectro-photometrically at 750 nm. The measurements are indicative of the quantity of phosphate released in relation to a calibration curve of phosphate in the range of 0 - 1 mMol/l.

###### Phosphatase stain

Components with phosphatase activity were detected by isoelectric focusing using a general phosphatase stain. The gel was incubated with a solution of α-naphthylphosphate and Fast Garnet GBC salt (Sigma, 0.1 & 0.2% (w/v), respectively) in 0.6 M sodium acetate buffer pH 5.5. The reaction, which results in the appearance of a black precipitate, was either terminated with methanol:acetic acid (30:10 %, v/v), or,

should the protein having phytase activity be further required, by rinsing with distilled water.

#### B. Purification of *A. ficuum* phytase

5

Phytase was purified to homogeneity from the culture broth of *A. ficuum* NRRL 3135. The broth was first made germ-free by filtration. The resulting culture filtrate was subsequently further concentrated in a Filtron ultrafiltration unit with 30 kD cutoff filters. The pH and ionic strength of the sample were adjusted for the purification procedure by washing the sample with 10 mM sodium acetate buffer pH 4.5. The final

10

concentration in this ultrafiltration procedure was approximately 20 fold. The sample was then applied to a cation exchanger (S-Sepharose Fast-Flow in a HR 16/10 20 ml column, both obtained from Pharmacia) in a Waters Preparative 650 Advanced Protein Purification System. The proteins bound were eluted with a sodium chloride gradient from 0-1 M in the sodium acetate buffer. Phytase eluted at approximately 250 mM NaCl. Phytase activity containing fractions were pooled, concentrated and desalted by ultrafiltration. The resulting solution was applied to an anion exchanger (Q-Sepharose Fast-Flow in a HR 16/10 20 ml column, Pharmacia), and the proteins were again eluted by a sodium chloride gradient from 0-1 M in the acetate buffer described above. Phytase was eluted from this column at approximately 200 mM NaCl.

15

The result of these purification steps is a partially purified phytase preparation with a specific activity of approximately 40-50 U/mg protein, indicating a 25-fold purification.

20

Analysis of the purity of the partially purified phytase indicated the presence of a major impurity with a molecular weight of approximately 100 kDa (Fig. 1B, sequence E). Isoelectric focusing indicated the presence of a number of phosphatase activity containing enzymes, including 3-4 phytase subforms (isoelectric points varying from 5.0-5.4) (Fig 1A, sequences A and B).

25

In order to obtain a homogeneous phytase preparation, a further two-fold purification was achieved by a subsequent separation of the components of the partially purified phytase by isoelectric focusing in a LKB Multiphor system on Ampholine PAG plates (pH range 4-6.5). The proteins with phosphatase activity (including the phytase) were detected by the general phosphatase staining procedure described above. The bands of interest were subsequently excised from the gel and the active protein was eluted by a 16 hr incubation of the gel slices in 10 mM sodium acetate buffer 5.5. The protein fractions were analysed in the specific phytase activity assay, as described in Example 2, thus discriminating the phytase fractions from other acid phosphatases. The final purification factor for phytase was approximately 60 fold (specific activity of final preparation 100 U/mg protein). In this final purification step it was also possible to isolate different subforms of phytase (Fig. 1A, sequences A and B).

30

35

Monoclonal antibodies directed against the *A. ficuum* phytase were prepared, providing an effective purification procedure. The antibody was coupled to cyanogen bromide-activated Sepharose 4B (5 mg/ml gel), and this matrix was used in an immunoaffinity column. The matrix was shown to bind approximately 1 mg phytase per ml. The phytase could be eluted from the affinity column with a pH 2.5 buffer (100 mM glycine-HCl, 500 mM NaCl) without any loss of activity. This procedure can be used to isolate homogeneous phytase from a crude culture filtrate in one single step with an 80% recovery and a 60-fold purification.

40

#### C. Deglycosylation of phytase

45

*A. ficuum* phytase (70 µg protein) was incubated with 2.5 U N-Glycanase (Genzyme) in 0.2 M sodium phosphate buffer pH 8.6 and 10 mM 1,10-phenanthroline in a total volume of 30 µl.

50

After 16 hrs at 37°C, the extent of deglycosylation was checked by electrophoresis (Phast System, Pharmacia). The apparent molecular weight of the phytase was found to decrease from 85 kDa to approximately 58.5 kDa. The periodic acid Schiff (PAS) sugar staining, which identifies native phytase as a glycoprotein, failed to detect any residual carbohydrates attached to the protein. The complete removal of carbohydrate was further substantiated by the sensitive lectin-blotting method. Native and deglycosylated phytase (both 1.5 µg) were run on a standard SDS-PAGE gel and electrophoretically transferred to a PVDF membrane (Immobilon, Millipore) in 25 mM TRIS-glycine buffer pH 8.3, 20% (v/v) methanol, for a period of 16 hrs at 30V.

55

The membrane was subsequently incubated with 1% (w/v) bovine serum albumin in phosphate buffered saline and incubated with concanavalin A-peroxidase (Sigma, 10 µg/ml in phosphate buffered saline). The peroxidase was then stained with 4-chloro-1-naphthol (Sigma).

This sensitive method also failed to detect any residual carbohydrate attached to the deglycosylated phytase.

After deglycosylation, phytase has completely lost its activity, possibly due to aggregation of the enzyme.

6

### Example 3

70

#### Determination of the amino acid sequence of phytase and design of oligonucleotide probes

##### A. Determination of the N-terminal amino acid sequence

15

Phytase was electrophoretically transferred from SDS-PAGE or from IEF-PAGE onto a PVDF blotting membrane (Immobilon, Millipore). Electroblotting was performed in 10 mM CAPS (3-cyclohexylamino-propanesulfonic acid) buffer pH 11.0, with 10% (v/v) methanol, for a period of 16 hrs. at 30V and 4 °C.

The protein was located with Coomassie Brilliant Blue staining. The band of interest was excised, further destained in methanol and subjected to gas-phase sequencing. The procedure has been carried out several times, using several individual preparations. The results obtained are given in Figure 1A (sequences A and B).

The amino acid sequence has also been determined for a 100 kDa protein that was present in crude preparations. The data obtained for this protein are given in Figure 1C. This sequence shows considerable homology with the acid phosphatase that has been isolated from *Aspergillus niger* (MacRae et al. (1988) Gene 71, 339-348).

30

##### B. Determination of internal amino acid sequences

##### Protein fragmentation by cyanogen bromide

Phytase, purified to homogeneity, was transferred into 100 mM NaHCO<sub>3</sub> by ultrafiltration (Microconcentrator Centricon 30, Amicon). The protein was subsequently lyophilized, dissolved in 70% trifluoroacetic acid (v/v), and incubated for 6 hr with an approximately 300-fold molar excess of CNBr. The reaction was terminated by dilution of the mixture with water. The resulting fragments were again lyophilized. The sample was then dissolved in SDS-PAGE sample buffer containing DTT (dithiothreitol), and the extent of fragmentation was determined by PAGE. Analytical PAGE was performed on a Pharmacia Phast-System unit, on 20% SDS-PAGE gels. The gels were prerun to create a continuous buffer system to improve the separation of the small peptides (according to the manual). Peptides were detected using a silver-staining technique known in the art, since Coomassie Brilliant Blue failed to detect the smallest peptide. The result of the procedure was a complete degradation of phytase into peptides with molecular weights of < 2.5 kDa, 36 kDa, 57 kDa and 80 kDa.

The peptides were isolated for gas-phase sequencing by SDS-Tricine-PAGE as described by Schagger & Jagow (1987) Anal. Biochem. 166, 368-379 followed by electroblotting as described above.

The N-terminus of the 57 kDa fragment is identical to the N-terminus of phytase as determined by Ullah (1988b, supra), with the exception of the first four amino acids which are absent (Figure 1A, sequence B). The N-terminal sequences of the 2.5 kDa and 36 kDa peptides are shown in Figure 1B as sequences A and

50 B.

##### C. Oligonucleotide probes

Oligonucleotide probes have been designed, based on the amino acid sequences given in Figure 1A and 1B, and were prepared using an Applied Biosystems ABI 380B DNA synthesizer. These oligonucleotides are given in Figure 2A and 2B.

55

Example 46 Hybridization of genomic blots and genomic libraries with a first set of oligonucleotide probes

Genomic DNA from *A. ficuum* has been isolated by grinding the mycelium in liquid nitrogen, using standard procedures (e.g. Yeiton et al (1984) Proc. Natl. Acad. Sci.-U.S.A., 1470-1474). A genomic library was constructed in the bacteriophage vector lambda EMBL3, using a partial *Sau* 3A digest of *A. ficuum* NRRL 3135 chromosomal DNA, according to standard techniques (e.g. Maniatis et al. (1982) Molecular cloning, a laboratory manual, Cold Spring Harbor Laboratory, New York). The thus-obtained genomic library contained 80 to 70 times the *A. ficuum* genome. The library was checked for the occurrence of plaques without insert by hybridization with the lambda EMBL3 stuffer fragment. Less than 1% of the plaques were observed to hybridize to the lambda EMBL3 probe. The insert size was 13 to 17 kb.

15 To identify conditions and probes that were suited for the screening of the genomic library, genomic DNA was digested with several restriction enzymes, separated on agarose gels and blotted onto Genescreen plus, using the manufacturers instructions. The blots were hybridized with all oligonucleotide probes. Hybridization was performed using conditions of varying stringency (6 x SSC, 40 to 60 °C for the hybridization; up to 0.2 x SSC, 65 °C for the washing). Probes 1068 and 1024 (Figure 2A) were selected for the screening of the genomic library, although no common DNA fragments could be identified that hybridized specifically with both probes. Acid-phosphatase probe 1025 (Figure 3) gave a specific and discrete hybridization signal and hence this probe was selected for screening the genomic library for the acid phosphatase gene.

Using all three probes, hybridizing plaques could be identified in the genomic library. The hybridization signal corresponding to probe 1025 (acid phosphatase) was strong and reproducible. Hybridization signals of variable intensity were observed using probes 1024 and 1068 (phytase). No cross hybridization between the two series was observed. All three series of plaques were rescreened and DNA was isolated from eight single, positive hybridizing plaques (Maniatis et al., supra). In each series, clones that contained identical hybridizing fragments could be identified, indicating that the inserts of said clones are related and probably overlap the same genomic DNA region. Again, no cross-hybridization could be demonstrated using the two phytase specific series (probes 1024 and 1068), indicating that, although both probes used to isolate the two series of clones were obtained from the N-terminal amino acid sequence of the protein, different genomic DNA fragments had been identified and cloned.

35 All three series of clones were hybridized with Northern blots containing mRNA isolated from induced and non-induced mycelium (Example 6). The acid phosphatase-specific clones, as well as the isolated internal 3.1 kb *Sall* fragment from these clones, hybridized exclusively to induced mRNA samples. The mRNA identified by the acid phosphatase-specific probes is about 1800 b in length, which agrees with the known size of the protein (68 kDa, Ullah and Cummins (1987) Prep. Biochem. 17, 397-422). No hybridization of the phytase-specific clones with specific mRNA's could be demonstrated. We have thus concluded that the above-described method was unsuccessful in cloning the gene encoding phytase. It may be further concluded that this failure is not due to a failure in the method used, since the method has been successfully applied to identify the gene encoding acid phosphatase. The lambda clone containing the acid phosphatase gene was deposited on April 24, 1989 at the Centraal Bureau voor Schimmelcultures, Baarn, The Netherlands and has been assigned accession number CBS 214.89. A 10 kb *Bam* HI fragment has been isolated from phage Z1 and subcloned into pUC19. This subclone contains the entire gene encoding acid phosphatase. The subclone, pAF 1-1 (Figure 5) was deposited on April 24, 1989 as CBS 213.89.

Example 5

50

Isolation of the gene encoding phytase, using a second set of oligonucleotide probes

65 Probes have been designed using the N-terminal amino acid sequence of CNBr-generated fragments (Figure 2B, probes 1295, 1296 and 1297) and have been hybridized with genomic DNA as described above. The feasibility of using these probes in the isolation of the gene encoding phytase was again studied by Southern hybridization of genomic blots with the probes. This time, hybridizing fragments of corresponding

lengths could be identified, using all three probes, despite the fact that the probes have been derived from non-overlapping regions. No hybridization was found between the new set of probes and the clones that have been isolated using the first set of probes (Example 4). Therefore, the genomic library was rescreened using all three probes in separate experiments. A subset of the clones (lambda AF201, 219, 241 and 243) isolated with each individual probe also hybridized with both other probes, indicating that in this case, using the three different probes, clones were isolated from a single genomic region. Attempts were made to hybridize the newly isolated clones with probes 1024 and 1068. In both cases, no hybridization with the newly isolated clones was observed under conditions in which both probes had successfully hybridized to the clones which were isolated using these probes (see Example 4). This demonstrates that the newly isolated clones have no homology to the probes derived from the N-terminus of the purified phytase.

A lambda EMBL3-clone, which hybridizes to all three probes (1285-1297), was named lambda AF201 (Figure 4) and was deposited on March 9, 1989 as CBS 153.89.

A 5.1 kb Bam HI fragment of lambda AF201 (subcloned in pUC19 and designated pAF 2-3, see Figure 4), hybridizing to all three oligonucleotide probes, was used to probe a Northern blot. In this case, a discrete mRNA having a size of 1800 bases was identified. This mRNA was found only in induced mycelium. Similar results were obtained when the oligonucleotides were used as probes. Therefore, using the new set of probes, a common DNA fragment has been identified, which hybridizes specifically to an induced mRNA. The length of this mRNA (1800 b) is sufficient to encode a protein of about 60 kDa, which is about the size of the non-glycosylated protein. Clearly, the isolated fragments contain at least part of the gene encoding phytase.

#### Example 6

25

#### Isolation of "induced" and "non-induced" mRNA

It is known from the literature that the synthesis of phytase by *A. ficuum* is subject to a stringent phosphate-dependent regulation (Han and Callagher (1987) J. Indust. Microbiol. 1, 295-301). Therefore, the demonstration that an isolated gene is subject to a similar regulation can be considered to support the evidence that the gene of interest has been cloned.

In order to isolate mRNA that has been synthesized under both producing and non-producing conditions, *A. ficuum* NRRL 3135 was grown as follows. Spores were first grown overnight in non-inducing medium. The next day, the mycelium was harvested, washed with sterile water and inoculated into either inducing or non-inducing medium. The medium used contains (per liter): 20 g corn starch; 7.5 g glucose; 0.5 g  $\text{MgSO}_4 \cdot 7 \text{H}_2\text{O}$ ; 0.2 g  $\text{FeSO}_4 \cdot 7 \text{H}_2\text{O}$ ; and 7.2 g  $\text{KNO}_3$ . For the induction of phytase, up to 2 g/l corn steep liquor was added to the medium, while non-inducing medium contains 2 g/l  $\text{K}_2\text{HPO}_4$ . The mycelium was grown for at least a further 100 hours. Samples were taken at selected intervals. Phytase production was followed by the phytase assay as described in Example 2A. Denatured mRNA was separated by electrophoresis and blotted onto Genescreen plus. The blots were hybridized with  $^{32}\text{P}$ -labelled pAF 2-3 or with the isolated 3.1 kb Sal I fragment from pAF 1-1 (acid phosphatase) from Example 4. The results are shown in Table 2.

Positive hybridization of the phytase specific 5.1 kb Bam HI fragment and the acid phosphatase specific 3.1 kb Sal I fragment with isolated mRNA is observed only when cells are grown under conditions which are known to induce the synthesis of phytase and acid phosphatases. From these results it has been concluded that the isolated genes are regulated as expected for phytase and acid phosphatases.

50

55

Table 2

Hybridization of Northern blots using the phytase-specific 5.1 kb BamHI fragment (A) or the acid phosphatase specific 3.1kb SalI fragment (B) as a probe; a + indicates the presence of the 1800 b phytase mRNA or the 1800 b acid phosphatase mRNA. The relative phytase activity was determined for the 24 hr. samples: induced cultures have 10 times more phytase activity than non-induced cultures.

		Time after inoculation	Induced	Non-induced
20	A	24 hours	+	-
	B	24 hours	+	-

Example 7

Evidence for the cloning of the phytase gene

To obtain definitive proof for the successful isolation of the gene encoding phytase, and to study the feasibility of increasing the expression of the cloned gene, the phytase gene was subcloned into a suitable vector and transformed to A. niger 402 (ATCC 9092). To this end, the phytase gene was isolated from the lambda clone AF201 as a 10 kb Nru I fragment and cloned into the Stu I site of the vector pAN 8-1 (Mattern, L.E. and Punt, P.J. (1988) Fungal Genetics Newsletter 35, 25) which contains the ble gene (conferring resistance to phleomycin) as a selection marker. The resulting construct was named pAF 28-1 (Figure 4) and was transformed to A. niger 402 according to the procedure as described in Example 9, with the exception that the protoplasts were plated on Aspergillus minimal medium supplemented with 30 µg phleomycin/ml and solidified with 0.75% agar. Single transformants were purified and isolated and were tested for production in shake flasks, as described in Examples 1 and 2. As controls, transformants possessing only the vector, as well as the untransformed host were also tested (Table 3). Only A. niger 402 containing pAF 28-1 appeared to produce a phytase that reacted with a specific monoclonal antibody directed against A. ficuum phytase. The phytase reacting with this monoclonal antibody could be eluted from an Immuno affinity column at pH 2.5 and was shown to be identical in molecular weight, degree of glycosylation, isoelectric point and specific activity to the A. ficuum phytase. This finding provides clear evidence that A. niger 402 cells transformed with pAF 28-1 express a phytase that is virtually identical to the A. ficuum phytase. Similar expression was not observed in either type of control cells.

Table 3

Strain	Phytase/Activity U/ml	% of phytase-activity adsorbed onto the immunoaffinity column
<u>A. niger</u> 402	0.5	0
<u>A. niger</u> 402 pAF 28-1	0.7	10
<u>A. niger</u> 402 pAN 8-1	0.5	0

Strains were grown under induced conditions (Example 6). Samples were taken after 96 hours of growth.

### Example 8

#### Characterization of the phytase gene.

The lambda clones containing the phytase gene have been analyzed by digestion with various restriction enzymes. A map of the genomic region encompassing the phytase gene is given in Figure 4. Defined restriction fragments have been subcloned in the cloning vector pUC19, as indicated in Figure 4.

It has previously been shown (Example 5) that the 5.1 kb Bam HI fragment present in pAF 2-3 encompasses at least part of the phytase gene. Moreover the oligonucleotide probes 1295 and 1297 (Figure 28) were shown to hybridize to the Sal I insert from pAF 2-7 (positions of pAF 2 clones are presented in Figure 4), while probe 1296 probably spans the Sal I site between the fragments in pAF 2-6 and pAF 2-7. The results of these experiments indicate that the phytase encoding sequence is located in the lefthand part of the Bam HI insert of pAF 2-3.

Subsequently the nucleotide sequences of the inserts of plasmids pAF 2-3, pAF 2-6, and pAF 2-7 have been determined completely using the dideoxy chain termination method (Sanger et al. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467) and shotgun strategies described by Messing et al. (1981, Nucl. Acids Res. 9, 309-321). In addition specific oligonucleotides were synthesized based on nucleotide sequence information obtained during the sequencing procedure.

The complete nucleotide sequence of clones pAF 2-3, pAF 2-6, and pAF 2-7 encompassing the chromosomal phytase gene locus is compiled in Figure 6, a graphic representation is given in Figure 7.

Analysis of the protein coding capacity of the complete sequence revealed that the N-terminal amino acid sequence of the mature protein was encoded starting from nucleotide position 381 (the N-terminus disclosed by Ullah is located at position 369). Furthermore, the N-terminal amino acid sequence of the 36 kDa and 2.5 kDa internal peptide fragments (see Figure 1B - sequences B and A) were found to be encoded at nucleotide positions 1101 and 1548, respectively. The open reading frame stops at nucleotide position 1713.

These findings clearly prove the identity of the characterized chromosomal locus as containing phytase encoding DNA sequence.

Directly upstream of the chromosomal sequence encoding the mature phytase protein, no ATG start codon can be found within the reading frame contiguous with the mature protein open reading frame; however, using intron-exon boundary characteristics, an intron can be postulated between nucleotide positions 254 and 355, bringing the ATG codon at nucleotide position 210 in frame with the mature phytase encoding open reading frame. The derived amino acid sequence of this N-terminal extension closely fits the rules for a secretion signal sequence as published by von Heyne (1983, Eur. J. Biochem. 133, 17-21).

To confirm these hypotheses the phytase cDNA was isolated by PCR-amplification with specific phytase primers and a total mRNA/cDNA population as template according to the procedures described below.

#### Isolation of poly A<sup>+</sup> RNA from Aspergillus ficuum.

Total RNA was isolated from A. ficuum NRRL 3135 grown under induced conditions as mentioned in



Example 6. Dry mycellum was frozen with liquid nitrogen and ground. Subsequently, the powder was homogenized in an Ultra-Turrax (full speed during 1 minute) in 3M LiCl, 8M urea at 0°C and maintained overnight at 4°C as described by Auffrey & Rougeon (Eur.J.Biochem., 107, 303-314, 1980). Total cellular RNA was obtained after centrifugation at 16,000 g for 30 minutes and two successive extractions with phenol: chloroform: isoamylalcohol (50:48:2). The RNA was precipitated with ethanol and dissolved in 1 ml 10 mM Tris-HCl (pH 7.4), 0.5% SDS. For poly A<sup>+</sup> selection the total RNA sample was heated for 5 minutes at 60°C, adjusted to 0.5 M NaCl and subsequently applied to an oligo(dT)-cellulose column. After several washes with a solution containing 10 mM Tris-HCl pH 7.4, 0.5% SDS and 0.1 M NaCl, the poly A<sup>+</sup> RNA was collected by elution with 10 mM Tris-HCl pH 7.4 and 0.5% SDS.

#### Preparation of the mRNA/cDNA complex

For the synthesis of the first cDNA strand 5 µg of poly A<sup>+</sup> RNA was dissolved in 18.5 µl H<sub>2</sub>O and the following components were added: 2.5 µl RNasin (30 U/µl); 10 µl of a buffer containing 50 mM Tris-HCl pH 7.6, 6 mM MgCl<sub>2</sub> and 40 mM KCl; 2 µl 1 M KCl; 5 µl 0.1 M DTT; 0.5 µl oligo (dT)<sub>12-18</sub> (2.5 mg/ml); 5 µl 8 mM dNTP-mix; 5 µl BSA (1 mg/ml) and 2.5 µl Moloney MLV reverse transcriptase (200 U/ml). The mixture was incubated for 30 minutes at 37°C and the reaction was stopped by addition of 10 µl 0.2 M EDTA and 50 µl H<sub>2</sub>O. An extraction was performed with chloroform and after centrifugation 110 µl 5 M NH<sub>4</sub>Ac and 440 µl ethanol were successively added to the supernatant. Precipitation of the mRNA/cDNA complex was performed in a dry ice/ethanol solution for 30 minutes. The mRNA/cDNA was collected by centrifugation, subsequently washed with 70% ice-cold ethanol and dissolved in 20 µl H<sub>2</sub>O.

#### Cloning of phytase cDNA fragments

Isolation of the cDNA-encoding phytase sequences were performed by the polymerase chain reaction (PCR) in two fragments. Four synthetic oligonucleotide primers were designed based on the genomic phytase sequence as presented in Figure 6.

Oligo 1: 5'-GGG.TAG.AAT.TCA.AAA.ATG.GGC.GTC.TCT.GCT.GTT.CTA-3'

Oligo 2: 5'-AGT.GAC.GAA.TTC.GTG.CTG.GTG.GAG.ATG.GTG.TCG-3'

Oligo 3: 5'-GAG.CAC.CAA.GCT.GAA.GGA.TCC-3'

Oligo 4: 5'-AAA.CTG.CAG.GCG.TTG.AGT.GTG.ATT.GTT.TAA.AGG.G-3'

Oligo 1 contains the nucleotide sequence downstream of the phytase ATG startcodon (position 210 to 231) flanked at the 5' border by an Eco RI-site; oligo 2 contains the nucleotide sequence immediately upstream of the Sal I-site (position 1129 to 1109) also flanked by an additional Eco RI-site; oligo 3 contains the nucleotide sequence around the Bam HI-site (position 845 to 865) and oligo 4 contains a nucleotide sequence positioned downstream of the phytase stopcodon (position 1890 to 1867) flanked by an additional Pst I-site.

The polymerase chain reactions were performed according to the supplier of Taq-polymerase (Cetus). As template the solution (1.5 µl) containing the mRNA/cDNA hybrids (described above) was used and as primers 0.3 µg of each of the oligos 1 and 2 in the reaction to amplify the N-terminal phytase cDNA part and oligos 3 and 4 in the reaction to amplify the C-terminal phytase cDNA part (see Figure 8). After denaturation (7 minutes at 100°C) and addition of 2 U Taq-polymerase the reaction mixtures were subjected to 25 amplification cycles (each: 2' at 55°C, 3' at 72°C, 1' at 94°C) in a DNA-amplifier of Perkin-Elmer/Cetus. In the last cycle the denaturation step was omitted. After digestion (Eco RI for the N-terminal cDNA part and Bam HI and Pst I for the C-terminal cDNA part), both cDNA fragments were cloned into the appropriate sites of pTZ18R (Promega).

The nucleotide sequence of both obtained PCR fragments was determined using the dideoxy chain termination technique (Sanger, supra) using synthetic oligonucleotides designed after the chromosomal phytase gene sequence, as primers and total amplified DNA as well as cloned cDNA fragments as template. The sequence of the cDNA region encoding the phytase protein and the derived amino acid sequence of the phytase protein are depicted in Figure 8.

The cDNA sequence confirmed the location of the intron postulated above, and indicated that no other introns were present within the chromosomal gene sequence.

The phytase gene encodes a primary translation product of 467 amino acids (MW 51091); processing of the primary translation product by cleaving off the signal peptide results in a mature phytase protein of 444 (MW 48851) or 448 (containing the first four N-terminal amino acids as published by Ullah, MW 49232)

amino acids.

### Example 9

#### Overexpression of phytase in *Aspergilli* by introduction of additional phytase genomic DNA copies

##### Construction expression vector pAF 2-2S

All constructs were made using standard molecular biological procedures, as described by Maniatis et al., (1982) Molecular cloning, A laboratory Manual, Cold Spring Harbor Laboratory, N.Y..

An expression vector pAF 2-2S was made by subcloning the 8 kb *Pvu* II DNA fragment of the phytase genomic clone lambda AF201, into the *Sma* I-site of pUC19. The derived plasmid was designated pAF 2-2 (Figure 4). As selection marker for the transformation to *Aspergillus*, the *Eco* RI/*Kpn* I DNA fragment of plasmid pGW325 (Wernars K. (1986), Thesis, Agriculture University, Wageningen, The Netherlands) containing the homologous *Aspergillus nidulans* *amdS* gene, was inserted into the *Eco* RI/*Kpn* I sites of pAF 2-2. The resulting expression vector was designated pAF 2-2S and is shown in Figure 9.

##### A. Overexpression of phytase in *A. ficuum* NRRL 3135.

The plasmid pAF 2-2S was introduced in *A. ficuum* NRRL 3135 using transformation procedures as described by Tilburn, J. et al. (1983) Gene 26, 205-221 and Kelly, J. & Hynes, M. (1985) EMBO J., 4, 475-479 with the following modifications:

- mycelium was grown on *Aspergillus* minimal medium (Cove, D. (1966) Biochem. Biophys. Acta, 113, 51-56) supplemented with 10 mM arginine and 10 mM proline for 16 hours at 30 °C in a rotary shaker at 300 rpm;

- only Novozym 234 (NOVO Industri), and no helicase, was used for formation of protoplasts;
- after 90 minutes of protoplast formation, 1 volume of STC buffer (1.2 M sorbitol, 10 mM Tris-HCl pH 7.5, 50 mM CaCl<sub>2</sub>) was added to the protoplast suspension and centrifuged at 2500 g at 4 °C for 10 minutes in a swinging-bucket rotor. The protoplasts were washed and resuspended in STC-buffer at a concentration of 10<sup>8</sup> cells/ml

- plasmid DNA was added in a volume of 10 µl in TE buffer (10 mM Tris-HCl pH 7.5, 0.1 mM EDTA) to 100 µl of the protoplast suspension;

- after incubation of the DNA-protoplast suspension at 0 °C for 25 minutes, 200 µl of PEG solution was added dropwise (25% PEG 4000 (Merck), 10 mM Tris-HCl pH 7.5, 50 mM CaCl<sub>2</sub>). Subsequently, 1 ml of PEG solution (60% PEG 4000 in 10 mM Tris-HCl pH 7.5, 50 mM CaCl<sub>2</sub>) was added slowly, with repeated mixing of the tubes. After incubation at room temperature, the suspensions were diluted with STC-buffer, mixed by inversion and centrifuged at 2000 g at 4 °C for 10 minutes. The protoplasts were resuspended gently in 200 µl STC-buffer and plated on *Aspergillus* minimal medium with 10 mM acetamide as the sole nitrogen source, 15 mM CsCl, 1 M sucrose, solidified with 0.75% bacteriological agar #1 (Oxoid). Growth was performed at 33 °C for 8-10 days.

Single transformants designated SP4, SP7 and SP8 were isolated, purified and tested for phytase production in shake flasks, using the process as described in Examples 1 and 2. As a control, transformants possessing only the vector (*amdS* gene in pUC19), as well as the untransformed host were tested.

Strains were grown under induced conditions (see Example 6) and samples were taken after 96 hours of growth. Analyses were performed by measuring the phytase activity (Table 4) and by isoelectric focusing polyacrylamide gelelectrophoresis (IEF-PAGE).

Samples of equal volume were taken from fermentations of *A. ficuum* and *A. ficuum* pAF 2-2S SP7, grown under identical conditions, and were applied onto an IEF-PAGE gel (pH-range 4.5-6, Phast-System, Pharmacia). The electrophoresis was performed according to the instructions of the manufacturer. Subsequently, the gels were either stained with the general protein stain Coomassie Brilliant Blue (Figure 10B), or with the general phosphatase activity staining described in Example 2 (Figure 10A).

A sample of *A. ficuum* phytase, purified to homogeneity (via immunoaffinity chromatography as described in Example 7), was also applied either alone, or mixed with a culture supernatant.

Phytase is present in the various samples in a number of isoforms (indicated with an asterisk), as has been mentioned in this invention. The two major isoenzymes are clearly visible in the purified phytase in lanes 3 and 4 with both staining procedures (A and B). The phytase bands are barely visible in the parent *A. ficuum* strain, and significantly increased in the pAF 2-2S SP7 transformant strain.

Table 4

Increase of phytase production by transformation of <i>A. ficuum</i> NRRL 3135.	
Strain	Phytase activity (U/ml)
<i>A. ficuum</i>	0.8
<i>A. ficuum</i> + control plasmid	0.8
<i>A. ficuum</i> pAF 2-2S SP8	7.6
<i>A. ficuum</i> pAF 2-2S SP7	6.7
<i>A. ficuum</i> pAF 2-2S SP4	4.3

#### B. Overexpression of phytase in *A. niger* CBS 513.88.

The expression vector pAF 2-2S was also introduced in *A. niger* CBS 513.88 by transformation procedures as described for *A. ficuum*. Single transformants were isolated, purified and tested for phytase production in shake flasks under induced growth conditions as described in Example 6.

Phytase expression levels of some transformants (designated as *A. niger* pAF 2-2S # 8, # 20 and # 33) and control strains were performed as described in Example 9A and are shown in Table 5.

*A. niger* transformants have phytase expression levels comparable with *A. ficuum* transformants. In addition this result indicates that the *A. ficuum* phytase promoter is active in *A. niger*.

Further analysis was performed on culture medium of transformant PAF 2-2S #8 by electrophoresis on an IEF-PAGE gel in the pH range of 4.5-6 on a Phast-System (Pharmacia) as described above. Equal volumes of the culture supernatants of the *A. niger* parent strain and of the transformant pAF 2-2S #8, grown under identical conditions, were applied onto the gel. The gels were run and subsequently stained as above.

The parent *A. niger* produces a very low amount of phytase, which could not be detected by gel electrophoresis. The strain PAF 2-2S #8 produces approx. 90 times more phytase, and this difference is clearly visible in Figure 11.

Several isoforms of the phytase enzyme are detected (indicated by asterisk). The general protein stain indicates that the intensity of the phytase protein bands is dramatically increased, while no other major protein bands appear.

Table 5

Phytase production by transformation of <i>A. niger</i> CBS 513.88 with pAF 2-2S.	
Strain	Phytase activity (U/ml)
<i>A. niger</i>	0.2
<i>A. niger</i> + control plasmid	0.2
<i>A. niger</i> pAF 2-2S # 8	14
<i>A. niger</i> pAF 2-2S # 33	5
<i>A. niger</i> pAF 2-2S # 20	4

Example 10

- 6 Phytase expression in *A. niger* transformed with expression vectors containing the *A. ficuum* phytase gene fused to the promoter and/or signal sequences of the *A. niger* amyloglucosidase (AG) gene.

Constructions of the expression vectors.

- 10 To obtain overexpression of phytase in *A. niger*, additional expression cassettes are derived in which the *A. ficuum* phytase gene is under control of the *A. niger* amyloglucosidase (AG) promoter in combination with different signal sequences. In p18FYT3 and p24FYT3 the respective 18 and 24 amino acid (aa) leader sequences of the AG gene from *A. niger* are fused to the phytase gene fragment encoding the mature protein. In the expression cassette pFYT3 the AG promoter sequence is fused to the phytase encoding sequence including the phytase leader sequence.

Construction of p18FYT3

- 20 Fusion of the AG-promoter and the 18 aa AG-leader sequence to the phytase sequence encoding the mature protein were performed by the Polymerase Chain Reaction method. In the PCR reactions two different templates were used: pAF 2-2S containing the entire phytase gene as described above and pAB6-1, a plasmid which contains the entire AG-locus from *A. niger*, which was isolated from a *A. niger* plasmid library, containing 13-15 kb *Hin* III fragments in pUC19. For the isolation, AG-specific oligos were used:
- 25 AG-1: 5'-GACAATGGCTACACCAGCACCGCAACGGACATTGTTTGGCCC-3'  
 AG-2: 5'-AAGCAGCCATTGCCCGAAGCCGAT-3'
- both based on the nucleotide sequence published for *A. niger* (Boel et al. (1984), EMBO J. 3, 1097-1102; Boel et al. (1984), Mol. and Cell. Biol. 4, 2306-2315). The oligonucleotide probes were derived from the sequence surrounding Intron 2: oligo AG-1 is located 3' of the intron and has a polarity identical to the AG mRNA and oligo AG-2 is found upstream of intron 2 and is chosen antiparallel to the AG mRNA. Plasmid pAB6-1 contains the AG gene on a 14.5 kb *Hin* III fragment (see Figure 12).

As primers for the PCR-amplifications four synthetic oligonucleotides were designed with the following sequence:

- 35 Oligo 1: 5'-CTCTGCAGGAATTCAGCTAG-3' (an AG-specific sequence around the *Eco*RI site approx. 250 bp upstream the ATG initiation codon).
- 40 Oligo 18-2: 5'-CGAGGCGGGGACTGCCAGTGCCAACCCTGTGCAGAC-3'  
 mature phytase <—> 18 AA AG-leader
- Oligo 18-3: 5'-GTCTGCACAGGGTTGSCACTGGCAGTCCCCGCCTCG-3'  
 18 aa AG-leader <—> mature phytase
- 45 Oligo 4: 5'-GGCACGAGGATCCTTCAGCTT-3' (a phytase specific sequence located at the *Bam*HI site on position 861)

- 50 The PCR was performed as described by Saiki et al. (1988), Science 239, 487-491, with minor modifications (see Example 8).

- To fuse the AG sequences to the phytase coding sequences two separate PCR's were carried out: the first reaction with pAB6-1 as template and oligos 1 and 18-2 as primers to amplify a 300 bp DNA fragment containing the 3'-part of the AG promoter and the 18 aa AG-leader sequence flanked at the 3'-border by the nucleotides of the phytase gene, and the second reaction with pAF 2-2S as template and oligos 18-3 and 4 as primers to amplify a 600 bp DNA fragment containing the 5' part of the phytase gene flanked at the 5'-

border by 18 nucleotides of the AG signal peptide. A schematic view of these amplifications is presented in Figure 13.

The two DNA fragments generated were purified by gelelectrophoresis and ethanol precipitation and used as templates in the third PCR with oligos 1 and 4 as primers to generate the AG-phytase fusion. The obtained DNA fragment was digested with Eco R1 and Bam HI and subcloned into pTZ18R. The resulted fusion was sequenced and designated p18FYT1.

The remaining (3.5 kb) upstream region of the AG-promoter was obtained by digestion of pAB6-1 with Kpn I and partially with Eco R1 and ligated to the 1.1 Kb Eco R1/Bam HI fragment of p18FYT1 and subsequently cloned into the Kpn I/Bam HI sites of pTZ18R. Plasmid p18FYT2 thus obtained is shown in Figure 15.

An additional Hln dIII restriction site was introduced by insertion of the synthetic fragment:

```

5' AATTCAAGCTTG      3'
3'      GTTCGAACTTAA 5'

```

into the Eco RI-site (flanking the amd S-gene) of pAF 2-2S. The obtained plasmid was designated pAF 2-2SH (Figure 14) and is used as starting plasmid to exchange the phytase promoter sequences by the PCR AG-phytase fusion DNA fragments.

For the final construction, p18FYT2 and pAF 2-2SH were digested with Kpn I and partially with Bam HI. The 4.6 kb DNA fragment of p18FYT2 and the 11 kb DNA fragment of pAF 2-2SH were isolated and purified by gel electrophoresis, subsequently ligated and transferred to E. coli. The derived expression cassette was designated p18FYT3 (Figure 15).

#### Construction of p24FYT3

Fusion of the AG-promoter and the 24 aa AG leader sequence to the mature phytase encoding sequence was performed by PCR-amplification as described above for the construction for p18FYT3 with the exception of the primers used. Two new primers were synthesized with the following sequence:

Oligo 24-2: 5'-CGAGCCGGGGACTGCCAGGCGCTTGGAAATCACATT-3'

mature phytase <—|> 24 AA AG-leader

Oligo 24-3: 5'-AATGTGATTTCGAAGCGCCTGGCAGTCCCCGCCTCG-3'

24 aa AG-leader <—|> mature phytase

Two separate PCR's were carried out: the first reaction with pAB 6-1 as template and oligos 1 and 24-2 as primers to amplify a 318 bp DNA fragment containing the 3'-part of the AG promoter and the 24 aa AG leader sequence flanked at the 3'-border by 18 nucleotides of the phytase gene and the second reaction with pAF 2-2S as template and oligos 24-3 and 4 as primers to amplify a DNA fragment containing the 5'-part of the phytase gene flanked at the 5'-border by 18 nucleotides of the 24 aa AG leader. A schematic view of these amplifications is presented in Figure 13.

For the construction of the final expression cassette p24FYT3 via the intermediate plasmids p24FYT1 and p24FYT2, the same cloning pathway/procedure was used as described for p18FYT1 and p18FYT2 to derive the expression cassette p18FYT3 (Figure 15).

#### Construction of pFYT3

Fusion of the AG-promoter to the phytase gene (including the phytase leader) sequence was also performed by PCR-amplification as described above for the construction of p18FYT3 with the exception of the primers used. Two additional primers were generated with the following sequence:

oligo fyt-2: 5'-AACAGCAGAGACGCCCATTGCTGAGGTGTAATGATG-3'  
 phytase leader <—|> AG-promoter  
 5 oligo fyt-3: 5'-CATCATTACACCTCAGCAATGGGCGTCTCTGCTGTT-3'  
 AG-promoter <—|> phytase leader

Two separate PCR's were carried out: the first reaction with pAB 6-1 as template and oligos 1 and fyt-2  
 10 as primers to amplify a 282 bp DNA fragment containing the 3'-part of the AG promoter flanked at the 3'-  
 border by 18 nucleotides of the phytase leader and the second reaction with pAF 2-2S as template and  
 oligos fyt-3 and 4 as primers to amplify a DNA-fragment containing the 5'-part of the phytase gene  
 (including the phytase leader) and flanked at the 5'-border by 18 nucleotides of the AG-promoter. A  
 schematic view of these amplifications is presented in Figure 13.

15 For the construction of the final expression cassette pFYT3 along the intermediate plasmids pFYT1 and  
 pFYT2, the same cloning pathway/procedure was used as described for p18FYT1 and p18FYT2 to derive  
 the expression cassette p18FYT3 (Figure 15).

#### 20 Expression of the phytase gene under the control of the AG promoter in *A. niger*

*E. coli* sequences were removed from the phytase expression cassettes described above by *Hin* dIII  
 digestion. Afterwards, the *A. niger* strain CBS 513.88 (deposited October 10, 1988) was transformed with 10  
 µg DNA fragment by procedures as described in Example 9. Single *A. niger* transformants from each  
 25 expression cassette were isolated, and spores were streaked on selective acetamide-agar plates. Spores of  
 each transformant were collected from cells grown for 3 days at 37 °C on 0.4% potato-dextrose (Oxoid,  
 England) agar plates. Phytase production was tested in shake flasks under the following growth conditions:

Approximately  $1 \times 10^8$  spores were inoculated in 100 ml pre-culture medium containing (per liter): 1 g  
 KH<sub>2</sub>PO<sub>4</sub>; 30 g maltose; 5 g yeast-extract; 10 g casein-hydrolysate; 0.5 g MgSO<sub>4</sub> · 7H<sub>2</sub>O and 3 g Tween 80.  
 30 The pH was adjusted to 5.5.

After growing overnight at 34 °C in a rotary shaker, 1 ml of the growing culture was inoculated in a 100  
 ml main-culture containing (per liter): 2 g KH<sub>2</sub>PO<sub>4</sub>; 70 g maltodextrin (Maldex MDO<sub>3</sub>, Amylum); 12.5 g  
 yeast-extract; 25 g casein-hydrolysate; 2 g K<sub>2</sub>SO<sub>4</sub>; 0.5 g MgSO<sub>4</sub> · 7H<sub>2</sub>O; 0.03 g ZnCl<sub>2</sub>; 0.02 g CaCl<sub>2</sub>; 0.05 g  
 MnSO<sub>4</sub> · 4 H<sub>2</sub>O and FeSO<sub>4</sub>. The pH was adjusted to 5.8.

35 The mycelium was grown for at least 140 hours. Phytase production was measured as described in  
 Example 2. The production results of several, random transformants obtained from each expression  
 cassette are shown in Table 6.

40

45

50

55

Table 6

Phytase production of several A. niger CBS 513.88 strains transformed with plasmids containing the A. ficuum phytase gene under control of the A. niger AG-promoter in combination with different leader sequences.

Expression cassette	Transformant #	Phytase activity (U/ml)
p18FYT3 (AG-promoter/ 18 aa AG-leader)	p18FYT3 # 240	82
	p18FYT3 # 242	84
	p18FYT3 # 243	62
	p18FYT3 # 244	43
	p18FYT3 # 245	80
	p18FYT3 # 246	82
	p18FYT3 # 250	110
p24FYT3 (AG-promoter/ 24 aa AG-leader)	p24FYT3 # 256	8
	p24FYT3 # 257	30
	p24FYT3 # 258	13
	p24FYT3 # 259	33
	p24FYT3 # 260	17
	p24FYT3 # 261	28
	p24FYT3 # 262	18
	p24FYT3 # 265	12
pFYT3 (AG-promoter/ phytase leader)	pFYT3 # 205	50
	pFYT3 # 282	280
	pFYT3 # 299	96
	pFYT3 # 302	220
	pFYT3 # 303	175
	pFYT3 # 304	150
	pFYT3 # 305	150
	pFYT3 # 312	140

The data clearly show high phytase expression levels in A. niger transformants containing the phytase gene under the control of the A. niger AG promoter. The data also show that the highest phytase production is obtained with the pFYT3 expression vector, which contains the phytase leader sequence. Similar expression vectors containing an intronless phytase gene after transformation to A. niger, resulted in phytase expression levels comparable to pFYT3 transformants of A. niger.

In addition, electrophoresis on an IEF-PAGE gel in the pH-range of 4.5-6 was performed on culture supernatants of transformants pFYT3 #205 and #282. Equal volumes of the culture supernatants of the A. niger parent strain and of both transformants, grown under identical conditions, were applied onto the gel, run and subsequently stained as described in Example 9. The parent A. niger produces a very low amount of phytase, which is not detected in this experiment. The strains pFYT3 #205 and #282 produce approx. 250 and 1400 times more phytase (compare phytase levels in Tables 4 and 5), and this difference is clearly visible in Figure 11. Several isoforms of the phytase enzyme are detected (indicated by an asterisk). The general protein stain indicates that the intensity of the phytase protein bands is dramatically increased, while no other major protein bands appear.

Overexpression of phytase in *A. ficuum* and *A. niger* grown on an industrial scaleA. *A. ficuum*

Strain *A. ficuum* pAF 2-25 #4 and *A. ficuum* NRRL 3135 were grown as described in Example 1. The transformant produced approximately 50 times more phytase as compared to the wild-type strain.

Table 7

Overexpression of phytase by a transformant of <i>A. ficuum</i> containing multiple phytase genes. Cells were grown as described in Example 1.		
Hours after inoculation	Phytase activity <i>A. ficuum</i> NRRL 3135	(U/ml Fermentation broth) <i>A. ficuum</i> pAF 2-2S #4
0	0	0
24	0	0
92	2	142
141	5	270

B. *A. niger*

Strain *A. niger* pAF 2-2S #8, a transformant of *A. niger* strain CBS 513.88 and the parent *A. niger* strain itself were grown as described in Example 1. The transformant produced approximately 1000 times more phytase as compared to the original *A. niger* parent strain (Table 8).

Table 8

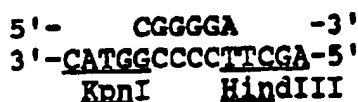
Overexpression of phytase by a transformant of <i>A. niger</i> (CBS 513.88) containing multiple phytase genes. Cells were grown as described in Example 1.		
Hours after inoculation	Phytase activity <i>A. niger</i> CBS 513.88	(U/ml fermentation broth) <i>A. niger</i> pAF 2.2 #8
0	0	0
24	0	5
92	0.1	65
141	0.1	95

Example 12

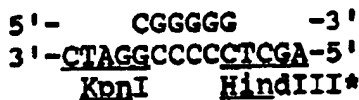
To construct the vector pREPFYT3, with which simultaneously phytase expression and AG gene replacement is achieved, pFYT3 is digested with Kpn I. With the obtained linear Kpn I DNA fragment, two separate ligations are performed.

Ligation 1 with the Kpn I-Hln dIII adaptor:





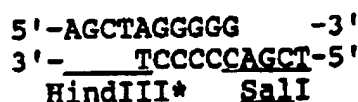
Ligation 2 with the Kpn I-Hin dIII\* adaptor, in which the Hin dIII restriction site will not restore after ligation:



Subsequently, ligation 1 is partially digested with Hin dIII. After removal of the amd S containing fragment by gel electrophoresis, the remaining DNA fragment is recircularized by ligation and transferred to E. coli. The obtained plasmid is denoted pFYT3ΔamdS (see Figure 16).

Ligation 2 is also digested with Hin dIII and the 4 kb DNA Hin dIII/Hin dIII\* fragment, containing the amd S gene, is isolated by gel electrophoresis, subsequently ligated to a partially Hin dIII digest of pFYT3ΔamdS and transferred to E. coli. The plasmid containing the amd S gene at the 3' end of the phytase gene is denoted pFYT3INT (see Figure 17).

To introduce the approx. 6 kb Sal I/Hin dIII DNA fragment of pAB6-1, containing the 3'-flanking AG sequence, pFYT3INT is partially digested with Hin dIII, ligated first to the adaptor:



(in which the Hin dIII\* restriction site will not restore after ligation) and subsequently with the Sal I/Hin dIII fragment of pAB6-1. After transformation to E. coli, the desired plasmid pREPFYT3, containing the 3' AG flanking sequence at the correct position, is obtained (Figure 18).

#### Expression of phytase in *A. niger* by AG gene replacement.

Before transformation of *A. niger* with pREPFYT3, the *E. coli* sequences in the plasmid are removed by Hin dIII digestion and gelelectrophoresis. The *A. niger* strain CBS 513.88 is transformed with 10 μg DNA fragment by procedures as described in Example 9. Selection and growth of transformants is performed as described in Example 9. Only a minority of the selected transformants lose AG activity (approx. 20%). Southern analysis of chromosomal DNA is performed on AG negative and phytase positive transformants to verify that the AG gene is indeed replaced by the phytase gene.

#### Example 13

#### Conservation of the phytase gene in different species.

To determine whether the phytase gene is highly conserved within microbial species, Southern analyses of chromosomal DNA from ten different species were performed with the *A. ficum* phytase cDNA as probe.

These chromosomal DNA analyses were performed on species from filamentous fungi, yeasts and bacteria. As an example, only a limited number from each group were chosen: for filamentous fungi, *Penicillium chrysogenum* and *Aspergillus niger*; for yeast, *Saccharomyces cerevisiae* and *Kluyveromyces lactis*; and for the procaryotic organisms the Gram-positive species, *Bacillus subtilis*, *Clostridium thermocellum*, and *Streptomyces lividans* and as an example for a gram-negative bacterium *Pseudomonas aeruginosa*.

High molecular weight chromosomal DNA from these species was digested with Pvu II and Bam HI

separately and subsequently electrophorized on a 0.7% agarose gel.

After transfer to nitrocellulose filters, the hybridization was performed overnight at low stringency (6 x SSC; 60 °C) with a <sup>32</sup>P-labeled 5'-phytase cDNA fragment (described in Example 8). Blots were washed in 6 x SSC at room temperature and exposed to X-ray for 18 hours.

As shown in Figures 19 a and b, discrete bands are observed in almost every lane, predicting a high degree of homology of the phytase gene between microbial species.

## Claims

1. A purified and isolated DNA sequence characterized in that said DNA sequence encodes a peptide or protein having phytase activity.
2. A purified and isolated DNA sequence according to Claim 1, further characterized in that said sequence is derived from a microbial source.
3. A purified and isolated DNA sequence according to Claim 1, further characterized in that said sequence is derived from a fungal source.
4. A purified and isolated DNA sequence according to Claim 1, further characterized in that said sequence is derived from an *Aspergillus* source.
5. A purified and isolated DNA sequence according to Claim 1, further characterized in that said sequence is derived from an *Aspergillus ficuum* or an *Aspergillus niger* source.
6. A purified and isolated DNA sequence according to Claim 1, further characterized in that said sequence encodes a phytase which exhibits the following characteristics:
  - a) provides a single band at 85 kDa on SDS-PAGE when expressed in an *Aspergillus* host;
  - b) has an apparent molecular weight after deglycosylation in the range of about 48 - 56.5 kDa;
  - c) has a specific activity of about 100 U/mg protein.
7. A purified and isolated DNA sequence characterized in that said sequence exhibits at least one of the following characteristics:
  - a) hybridizes to an oligonucleotide probe derived from the DNA sequence as disclosed in Figure 6;
  - b) hybridizes to an oligonucleotide probe derived from the cDNA sequence as disclosed in Figure 8.
8. An expression construct characterized in that a DNA sequence according any one of claims 1-7 is operably linked to a regulatory region capable of directing the expression of a protein or peptide having phytase activity in a suitable expression host.
9. The expression construct of Claim 8 further characterized in that the regulatory region also contains a secretory leader sequence providing for the secretion of the expressed protein or peptide having phytase activity.
10. The expression construct of Claim 9 characterized in that the AG promoter is used to direct the expression of the protein or peptide having phytase activity.
11. The expression construct of Claim 10 further characterized in that a homologous phytase leader sequence is used to provide for the secretion of the expressed protein or peptide having phytase activity.
12. The expression construct of Claim 10 further characterized in that the 18 amino acid AG leader sequence is used to provide for the secretion of the expressed protein or peptide having phytase activity.
13. The expression construct of Claim 10 further characterized in that the 24 amino acid AG leader sequence is used to provide for the secretion of the expressed protein or peptide having phytase activity.
14. The expression construct of Claim 9 characterized in that a homologous phytase promoter is used to direct the expression of the protein or peptide having phytase activity.
15. The expression construct of Claim 14 further characterized in that a homologous phytase leader sequence is used to provide for the secretion of the expressed protein or peptide having phytase activity.
16. The expression construct of Claim 14 further characterized in that the 18 amino acid AG leader sequence is used to provide for the secretion of the expressed protein or peptide having phytase activity.
17. The expression construct of Claim 14 further characterized in that the 24 amino acid AG leader sequence is used to provide for the secretion of the expressed protein or peptide having phytase activity.
18. A vector capable of transforming a host cell characterized in that said vector contains an expression construct according to any one of Claims 8 to 17.
19. A vector according to Claim 18, further characterized in that said vector is a plasmid.
20. A vector according to Claim 18, further characterized in that said vector is a plasmid selected from the group consisting of pAF 28-1, pAF 2-2S, pAF 2-2, pAF 2-3, pAF 2-4, pAF 2-6, pAF 2-7, p18FYT3, p24FYT3 and pFYT3.
21. A transformed host cell characterized in that said host cell is transformed with a vector according to any

one of Claims 18 to 20.

22. A transformed host cell according to Claim 21, which is selected from the group consisting of bacteria, yeasts and fungi.

23. A transformed host cell according to Claim 22, which is selected from the group consisting of Aspergillus, Trichoderma, Penicillium, Mucor, Bacillus, Kluyveromyces and Saccharomyces.

24. A transformed host cell according to Claim 22, which is selected from the group consisting of Aspergillus niger, Aspergillus ficuum, Aspergillus awamori, Aspergillus oryzae, Trichoderma reesei, Mucor miehei, Kluyveromyces lactis, Saccharomyces cerevisiae, Bacillus subtilis and Bacillus licheniformis.

25. A process for the production of a peptide or protein having phytase activity, characterized in that a transformed host cell according to any one of Claims 21 to 24 is cultured under conditions conducive to the production of said peptide or protein having phytase activity.

26. A peptide or protein having phytase activity, characterized in that said peptide or protein having phytase activity is produced by a process according to Claim 25.

27. A phytase characterized in that the phytase exhibits the following characteristics:

- a) provides a single band at 85 kDa on SDS-PAGE when expressed in an Aspergillus host;
- b) has an apparent molecular weight after deglycosylation in the range of about 48 - 56.5 kDa;
- c) has a specific activity of about 100 U/mg protein.

28. A feed for animals characterized in that said feed contains a peptide or protein having phytase activity according to either of Claims 26 and 27.

29. Use of a peptide or protein having phytase activity according to either of Claims 26 and 27 for the conversion of phytate to inositol and inorganic phosphate.

30. A process for promoting the growth of animals characterized in that an animal is fed a diet which is comprised of a feedstuff supplemented with a phytase according to either of claims 26 and 27.

31. A process for the reduction of levels of phytate in animal manure characterized in that an animal is fed a diet which is comprised of a feedstuff supplemented with a phytase according to either of claims 26 and 27 in an amount effective in converting phytate contained in the feedstuff to inositol and inorganic phosphate.

30

35

40

45

50

55

## N-TERMINAL AMINOACID SEQUENCES

A

B

C

## POSITION

01			LEU
02			ALA
03			VAL
04			PRO
05		ALA	ALA
06		SER	SER
07		---	ARG
08	----**	---	ASN
09	GLN	GLN	GLN
10	SER	SER	SER
11	SER	SER	SER
12	---	---	GLY
13	ASP	ASP	ASP
14	THR	THR	THR
15	VAL	VAL	VAL
16	ASP	ASP	ASP
17	GLN	GLN	
18		GLY	
19		TYR	
20		GLN	
21		ARG	
22		PHE	
23		SER	
24		GLU	
25		THR	
26		SER	
27		HIS	
28		LEU	
29		ARG	
30		(GLY)*	
31		GLN	
32		TYR	
33		ALA	
34		PRO	
35		PHE	
36		PHE	
37		(ASP)	
38		LEU	
39		ALA	

Fig. 1A

## PEPTIDE AMINOACID SEQUENCES

A B C D E

## POSITION

01	GLN	(TRP)*	MET	ALA	VAL
02	-----	SER*	MET	SER	VAL
03	GLN	PHE	GLN	SER	ASP
04	ALA	ASP	CYS	ALA	---
05	GLU	THR	GLN	GLU	ARG
06	GLN	ILE	ALA	LYS	PHE
07	GLU	SER	GLU	GLY	PRO
08	PRO	THR	GLN	TYR	TYR
09	LEU	SER	GLU	ASP	THR
10	VAL	THR	PRO	LEU	GLY
11	(ARG)	VAL	LEU	VAL	---
12	VAL	ASP	VAL	VAL	ALA
13	LEU	THR	ARG		
14	VAL	LYS	VAL		
15	ASN	LEU	LEU		
16	(ASP)	SER	VAL		
17	(ARG)	PRO	ASN		
18	(VAL)	PHE	ASP		
19	VAL	(CYS)	ARG		
20	PRO	(ASP)			
21		LEU			
22		PHE			
23		THR			

FIG. 1B

## N-TERMINUS 100KD PROTEIN

### POSITION

01	VAL
02	VAL
03	ASP
04	GLU
05	ARG
06	PHE
07	PRO
08	TYR
09	THR
10	GLY

FIG. 1C



[illegible]

AB1295 : GTC.CGC.CTC.GTC.CTC.CGC.GAG.CA  
T G T T T C A C

Peptide B : 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26  
(Trp)-Ser-Phe-Asp-Thr-Ile-Ser-Thr-Ser-Thr-Val-Asp-Thr-Lys-Leu-Ser-Pro-Phe-(Cys)-(Asp)-Leu-Phe-Thr-(Thr)-(Asp)-(Glu)

(TTGG)-TCG-TTT-GAT-ACG-ATA-TCG-ACG-TGG-GAT-ACG-GTG-GAT-ACG-CTG-TGG-CCG-TTT-(TGT)-(GAT)-CTG-TTT-ACG-(ACG)-(GAT)-(GAG)
A C C A T A A A A A A A A C C C A C A C A
T T C T Y T Y T T Y T Y T T T T T T T
C C C C C C C C C C C C C C C C C C
AGT AGT AGT TGG AGT TGG AGT TGG
C C C C C A C A

AB1296 : 3'-AAG.CTG.TGC.TAG.AGG.TGG.AGG.TGG.CAC.CTG.TGC.TTC  
TCC C TCC C

AB1297 : 3'-GGC.AAG.(ACG).(CTG).GAG.AAG.TGC.(TGC).(CTG).(CTC)-  
G C G G

	27	28	29	30	31	32	33
(Cys)-	(His)	(Tyr)	(Arg/Gly)	(Tyr)	(Leu)	(Tyr)	(Leu)
(Glu)	(Asn)	(Gln)	(Ser)	(Ala)	(Val)	(Ile)	(Met)
(Lys)	(Phe)	(Thr)	(Pro)	(Glu)	(Trp)	(His)	(Glu)
(Gly)	(Ala)	(Acg/Aat)	(Tat)	(CGG/ggg)	(TAT)	-CTG	
C	T	A	C	C	A	C	A
	C	T			T	T	T
		C			C	C	C
					AGG		TTG
					A		A

**(ACG).(TAG).(T**

**Figure 2B**





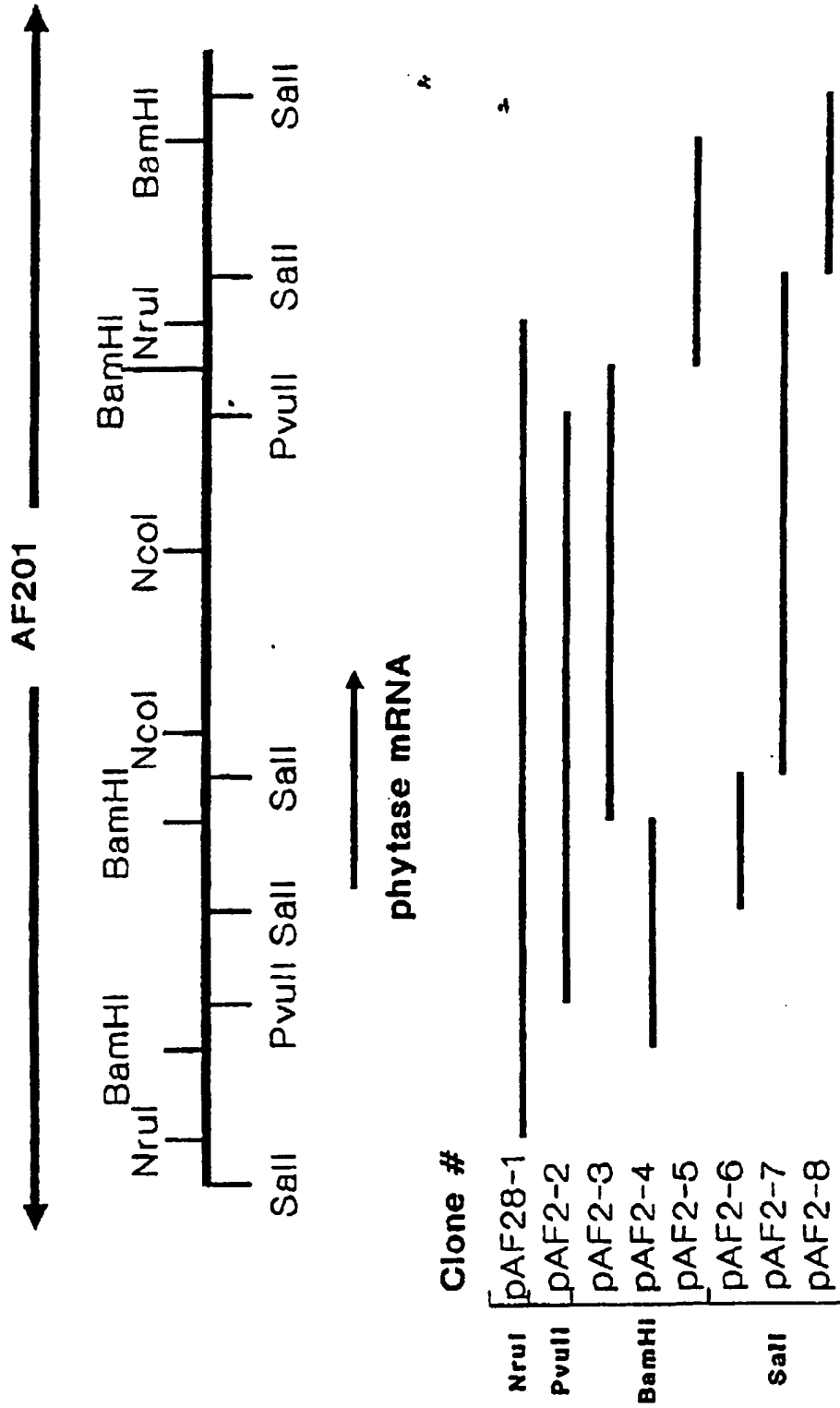


Figure 4

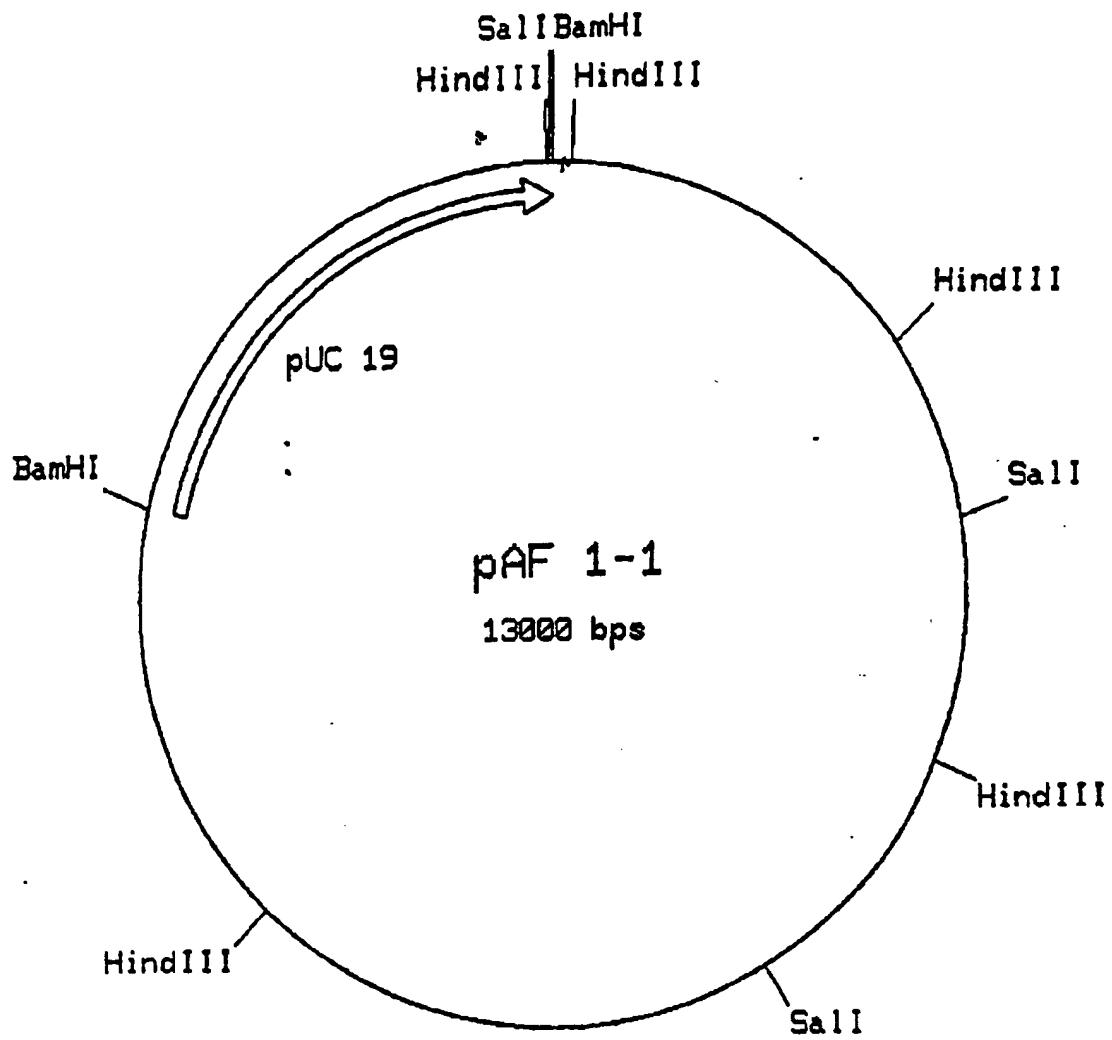


Figure 5

Figure 6

1 GTCGACTTCCCGTCCTATTCCGGCCTCGTCCGCTGAAGATCCATCCCACCATTGCACGTGG  
SalI

61 GCCACCTTTGTGAGCTTCTAACCTGAACTGGTAGAGTATCACACACCATGCCAAGGTGGG

121 ATGAAGGGGTTATATGAGACCGTCCGGTCCGGCGCGATGGCCGTAGCTGCCACTCGCTGC

181 TGTGCAAGAAATTACTTCTCATAGGCATCATGGGCGTCTCTGCTGTTCTACTTCCTTTGT  
translation start

241 ATCTCCTGTCTGGGTATGCTAAGCACCACAATCAAAGTCTAATAAGGACCGTCCCTTCCG  
intron start

301 AGGGCCCCTGAAGCTCGGACTGTGTGGGACTACTGATCGCTGACTATCTGTGCAGAGTCA  
intron end

361 CCTCCGGACTGGCAGTCCCCGCCTCGAGAAATCAATCCAGTTGCGATACGGTCGATCAGG

421 GGTATCAATGCTTCTCCGAGACTTCGCATCTTTGGGGTCAATACGCACCGTTCTTCTCTC

481 TGGCAAACGAATCGGTCATCTCCCTGAGGTGCCCGCCGGATGCAGAGTCACTTTCCGTC

541 AGGTCTCTCCCGTCATGGAGCGCGTATCCGACCGACTCCAAGGGCAAGAAATACTCCG

601 CTCTCATTGAGGAGATCCAGCAGAACGCGACCACCTTTGACGGAAAAATATGCCTTCCTGA

661 AGACATACAACTACAGCTTGGGTGCAGATGACCTGACTCCCTTCGGAGAACAGGAGCTAG

721 TCAACTCCGGCATCAAGTTCTACCAGCGGTACGAATCGCTCACAAGGAACATCGTTCCAT

781 TCATCCGATCCTCTGGCTCCAGCCGCGTGATCCCTCCGGCAAGAAATTCATCGAGGGCT

841 TCCAGAGCACCAAGCTGAAGGATCCTCGTGCCAGCCCGGCCAATCGTCGCCCAAGATCG  
BamHI

901 ACGTGGTCATTTCCCTAGCCAGCTCATCCAACAACACTCTCGACCCAGGCACCTGCACTG

961 TCTTCGAAGACAGCQAATTGGCCGATACCGTCGAAGCCAATTTACCGCCACGTTCTGTC

1021 CCTCCATTCTGTCAACGTCTGGAGAACGACCTGTCCGGTGTGACTCTCACAGACACAGAAG

Figure 6 (continued 1)

1081 TGACCTACCTCATGGACATGTGCTCCTTCGACACCATCTCCACCAGCACCCGTGGACACCA  
SalI

1141 AGCTGTCCCCCTTCTGTGACCTGTTACCCATGACGAATGGATCAACTACGACTACCTCC

1201 AGTCCTTGAAAAAGTATTACGGCCATGGTGCAGGTAACCCGCTCGGCCCGACCCAGGGCG

1261 TCGGCTACGCTAACGAGCTCATCGCCCGTCTGACCCACTCGCCTGTCCACGATGACACCA

1321 GTTCCAACCACACTTTGGAATCGAGCCCGGTACCTTTCCGCTCAACTCTACTCTCTACG

1381 CGGACTTTTCGCATGACAACGGCATCATCTCCATTCTCTTTGCTTTAGGTCTGTACAACG

1441 GCACTAAGCCGCTATCTACCACGACCGTGGAGAATATCACCCAGACAGATGGATTCTCGT

1501 CTGCTTGGACCGTTCCGTTTGCTTCGCGTTTGTACGTCGAGATGATGCAGTGTGAGGCGG

1561 AGCAGGAGCCGCTGGTCCGTGTCTTGGTTAATGATCGCGTTGTCCCGCTGCATGGGTGTC

1621 CGGTTGATGCTTTGGGGAGATGTACCCGGGATAGCTTTGTGAGGGGGTTGAGCTTTGCTA

1681 GATCTGGGGGTGATTGGGCGGAGTGTCTTCTTAGCTGAATTACCTTGATGAATGGTATG  
translation stop

1741 TATCACATTGCATATCATTAGCACTTCAGGTATGTATTATCGAAGATGTATATCGAAAGG

1801 ATCAATGOTGACTGTCACTGGTTATCTGAATATCCCTCTATACCTCGTCCACAAACCAAT

1861 CATCACCTTTTAAACAATCACACTCAACGCACAGCGTACAAACGAACAAACGCACAAAGA

1921 ATATTTTACACTCCTCCCCAACGCAATACCAACCGCAATTCATCATACCTCATATAAATA

1981 CAATACAATACAATATCTCCATCCCTACCCCTCAAGTCCACCCATCCTATAATCAATCCCT

2041 ACTTACTTACTTCTCCCCCTCCCCCTCACCCCTCCGAGAACTACCCCCGAAGTAGTAAT

2101 AGTAGTAGTAGAAGAAGCAGACGACCTCTCCACCAATCTCTTCGGCCTCTTATCCCCATA

Figure 6 (continued 2)

2161 CGCTACACAAAACCCCCACCCCGTTAGCATGCACTCAGAAAATAATCAAAAATAACTAAG  
2221 AAGGAAAAAAGAAGAAAGAAAGGTTACATACTGCTCTCATACAACTCCAAGACGTATA  
2281 CATCAAGATGGGCAATCCCACTTACTGATATCCATCTATGAACCCATTCCCATCCCAC  
2341 GTTAGTTGATTACTTTACTTAGAAGAAGAAAAAGGGAAGGGAAGCGAAAGAACTGGATGG  
2401 GATTGAGTTAGTGCTCACCGTCTCGCAGCAAGTTTATATTCTTTTGTGTTGGCGGATATCT  
2461 TTCACTGCTCTGCTGGACGTTGTCACGGGGTGGTAGTGGTTGGCGGTGGTGAGGGTCCA  
2521 TGATCACTCTTGGTTTGGGGGGTTGTTGTTGTCGTTGTTGTTGTTGTTGGGTGGGCATTT  
2581 TCTTTTCTTCACTTGGGGATTATTATTTGGAATTGGTTAGTTTGAGTGAGTGGGTAATAT  
2641 TGAATGGGTGATTATTGGGAATGAAGTAGATTTGGCTATGAATGGTTGATGGGATGGAAT  
2701 GAATGGATGGATGAATAGATGGAGCGGAAAAAGTCAGGTGGTTTGAGGTTCCGATTATTA  
2761 TCTTTGTGCTGAGGCATCACTCTCCATCTATGTTGTTCTTTCTATACCGATCTACCAGA  
2821 GCTAAGTTGACTGATTCTACCACAGTGCACAATAAGTATGTACTTATTTCAATTTAGAGTA  
2881 TTTAGATTAAACCGCTGTGCTATTTGCCGTAGCTTTCCACCCAATTTGGAAGTTGGAAGA  
2941 ATTAAAACTCATCCTACAGTACAGAATAGAAGTAAAAGGAGAAGAGAAAAACAAGATAAT  
3001 ACAACCAGTCCAGGTCCATTCTAGATCTCGAATGACCACCAAATAAGAAAGCAACAAGCA  
3061 AGTAAGCAAAGCATTTGTTCTAAATGAACGCCAATAACTTCATCGCCTGCCTTTGAAACTG  
3121 AACGCTATGCACGAATGGCTCGAAATGATTCCTTAACCTCCGTAGTATTGAGAGTGAGAG  
3181 GAAAAGAAAAAAGAGACAGAAAAGCTGACCATGGGAAAGAACATGATCAGTCGGGAAT

Figure 6 (continued 3)

3241 GGATCTGCGGGTTGAGATAGATATGAGTTGCCTCGCAGATCCGGTGACAAGATAAGAGAA  
3301 TTGGGAGATGTGATCAGCCACTGTAAC TTCATCAAGCATCGACATTCAACGGTCGGGTCT  
3361 GCGGGTTGAGATGCAAGTTGAGATGCCACGCAGACCCGAACAGACTGAGAGATGTGAGAC  
3421 TTTTGAACCACTGTGACTTCATCAAGCATCAAAACACACTCCATGGTCAATCGGTTAGGG  
3481 TGTGAGGGTTGATATGCCAGGTTGATGCCACGCAGACCCGAACCGACTGAGAAATATGA  
3541 AAAGTTGGACAGCCACTTCATCTTCATCAAGCGTAAAACCCCAATCAATGGTAAATCGAA  
3601 AACGAATCTGCGGGCTGATGTGAAATGAGACGAATGCCTCGCAGATTGGAAGACACGTA  
3661 AATCGAGATGAACAATCACTTTAACTTCATCAAAGCCTTAAATCACCCAATGGCCAGTCT  
3721 ATTGCGGTCTGCGGGTTGAGGTTCTGTTGAGATGCCACGCAGACTGCGAACATGCGATG  
3781 CATTATAAGTTGGACGAGTGTAGACTGACCATTGATAACCGAGATAAACAATCACTTCAA  
3841 CTTTCATCAAAGCCTTAAATCACTCAATGGCCAGTCTGTTTGCGGTCTGCGGGCTGATACC  
3901 CAAGTTGCGATGCCACGCAGACTGCAAACATTGATCGAGAGACGAGAAAAACAACGCACT  
3961 TTAAC TTCAACAAAAGCCTTTCAATCAGTCAATGGCCAGTCTGTTGCGGTCTGCGGGCT  
4021 GATATGCGAGTTGAGGTGCCTCGCAGACCCGGAACATGCGATGTAATTTCTTAGTTAGAC  
4081 GAGTGCCTGGCCATTGAGAAACGAGAGAAACAACCACTTTAACTTCATGAAAGCCTTGAA  
4141 CTA CTCAATGACCCCTCTGTTGCGGTCTGCGGGCTGATATTCGAGTTGAGATGCCACGC  
4201 AGACCGCCAACATGCCATGTATCATGTAAGTTAGATGAGTGA CTGGCCATTGAGAAACGA  
4261 GAGAAACAACCACACTTCATGAGAGCCTTAAATTATTCAATGACCAGTCTGTTACGGTC

Figure 6 (continued 4)

4321 TGGGGTTGGTATGCCAGTCGAGGTCCTCGCAGACCGGAACATGCGATGTTTTCGATG  
4381 GACGAGTGAAGCCTGACGATCGAGAACTATCTCAGTTGGGTTGGCCATTGCGCTGGCCGT  
4441 TGGGTTTAGTATTAGGATCGTCAGGTTTGTCCGATGGAACGTTCCGTTTGGCTGCCGTTGG  
4501 CGCGACGAGCCCTCTCCTCGGCCTGATTCTGAAATTCTGCAATCAGGGCAGCCGCAGCAC  
4561 GCGGACGGGACGTCCTCCAGGAGCTGTGTTGAAGTTTCGGGGTGGCGGTCCAGAAGGGGG  
4621 AGTTACATTAAAAGCCTCATAGATGTCTTTGGGTGGTTCCGGGGGGCCCATCGCAAGATC  
4681 TTCTGGAGTTGTGCGTCTGATCATCTCTTQAGTGAATTGCGACGCAGACCGAGCTTCAG  
4741 GATTTTGAAGGGCTGGATCGCTCCTGCTGACTCTTTCCTCAGCGGGCTTCGTCTCGGC  
4801 AGTCTTCATTTCCGGGGCTGATCTTCCATCTCAGAATGGGATCGCTTTCTGCTCGCTGC  
4861 ACCCGCTCCTCCCTTCAAGGTCAGCTTGATGCCAGCGTCTTGGGCGGCTCAGCTGGTGG  
4921 AGTTGGTTCCGGCTCTGGCTCCCTCCGGCGTGGCTTGGGCACTTGAGTAGTCTCTGAGGC  
4981 TTCGCCCGCGCGCGCTTTGCGAGTCGGCTCCTTGGTCTCTTTGGCCTCTTCACTTCACC  
5041 TGGACCGTCTTTCCGGGCGCTTTCATCGTGCTGAGCGATCAAGGTTTGGATGTAGGCAGC  
5101 CGGCATCATTCGATCAACGGCAATTCCTCTCTTGGGGGCTCCTCCCGAGCCTTGATTGT  
5161 CGCCTTGACCTCGTCCACGTTTTTCGAAGAAGAAAGGCATCTTGTATCCTGAGGCAAGTT  
5221 GCGCTCTCCCATGCGTGGGGATATCCGAAGATGCGGTCTTCTCGAACTGTTTCATGAGAC  
5281 TTCAGACGAATTGGAGGCTGGGGGAGCAATTTGTCTCCGTAGGTGTTGTTAGGGCGGAAC  
5341 CAAQAATAGCCTTCGCCTACAACGACAAGCTCTTCGCCAAATTTATTTTTTTGGCCTGTA



Figure 6 (continued 5)

5401 AAAACGAACCCATCCTCGTCAGTCCACCGGTGCGTCTCGGACGTAGAGATTGGCTTACTT  
5461 ATTCCCTCAACCCGATCTCTGCGTGGGCTGCGCTTCGGATGCGGCTCGGTACGGGT  
5521 CCGCCTCGGACTGCACCGCTGGAGTTTCGGTCTTCTTCTCCTGCTTCTCCAGGTACTCCT  
5581 TCGGTAAGTCTTCGATCAGCCTCGGCTTCGGATGACTGCTCAAATTCTGGAGCAACAGCT  
5641 GCGCGGCGCAGGTCAAGCAGGCGGTTTGCTAAAACTGCCCATTTTCCATCGACACCTGCC  
5701 TCGACGCGCTGTGCAAAACAGCTGTTTTCGCATTGGCCTGTTTGTGGCAGCGCTCTTC  
5761 TTGACTGCTGCCTTGCCCTTTACTTCCTTGAGAGCAGACTCTGGCTTAGATGATGGTGCA  
5821 CGGTTTCTGCGGAAGCGCGCTCAGATTCCAAAATTCCATAGCTTTAATGGTAGGCTTT  
5881 CTGGTTCTTCCAGAAAGTCCGCGCAGCTGACGTAGTGGTTGAGTAGCTGGCAGTTGGGAT  
5941 CCTGGGCGCTCATTGGAACCATCAAGACCAAATTTGTTCCATACATATCAGCATGGTAT  
6001 TCAAAAGGAAAATTTGCGCGTACGGAAGTACTGCGTTCCGATTCCGGGTGTATCCAAGTCG  
6061 TATCCAGACATGGTGTGGAATTCAGCCTTGCTGTCAAGAGCAGGGGTACTTTCAATGCTG  
6121 TCAGCAACCACCGGCCAAAGGGCGTCTTCGGGAAAGAAGGTGTTTCAAGAGAAGCGTCA  
6181 TCCACGGCCTGGCTTGCGGCGTTGATTGCAGACTTTCGAGTAGATCGCTGAGGTGCGGAA  
6241 CTGGTTCGAGTAGCAACCTGTGAATTGGCAGCCTTGTGACTGCTTCGATTCACTGCAGAG  
6301 ACGGAGTAGACTGCAGTGAATTTGGAATTCTGAGTCGCAGCCATTCTGCAATTTGCGTTGCG  
6361 CGCGACGAGATCTCGCAGTCGTGGTACGAGGAGTAGAGCGAGGCTGCGTAGCAGTGTTCG  
6421 AAGCTTGGTGCTAGCCTCCTGGGCTTCAGCAGCTTCAGCAGTGGTGGCAGACGCAGCAGA

Figure 6 (continued 6)

6481 ATTAGCGGAGCTTTATCGGCTTTGCCGCTCTGACCGTTGGGAGTAGAAGTGAGAGAAGAG

6541 GTAGAGTCCACGGAAGAAGTCTTCTCGCTGTTCTCAAAGCCGTTTCAGCTTTGCTGGCATA

6601 GACTTACCGCTCTTGCCGGCTGTTGGAAGCGGAAGAGTTCATGGCGGGAGAGGAGACGTTA

6661 GAAGTAGACATGGTGGGGTTTGTGACGGGTTTTGAGTAACAAGAGACTTGGCTCGATCT

6721 TTGAGTGTCTTGACAGAAAGTTATGCAACGTCGAC  
SalI

# phytase locus

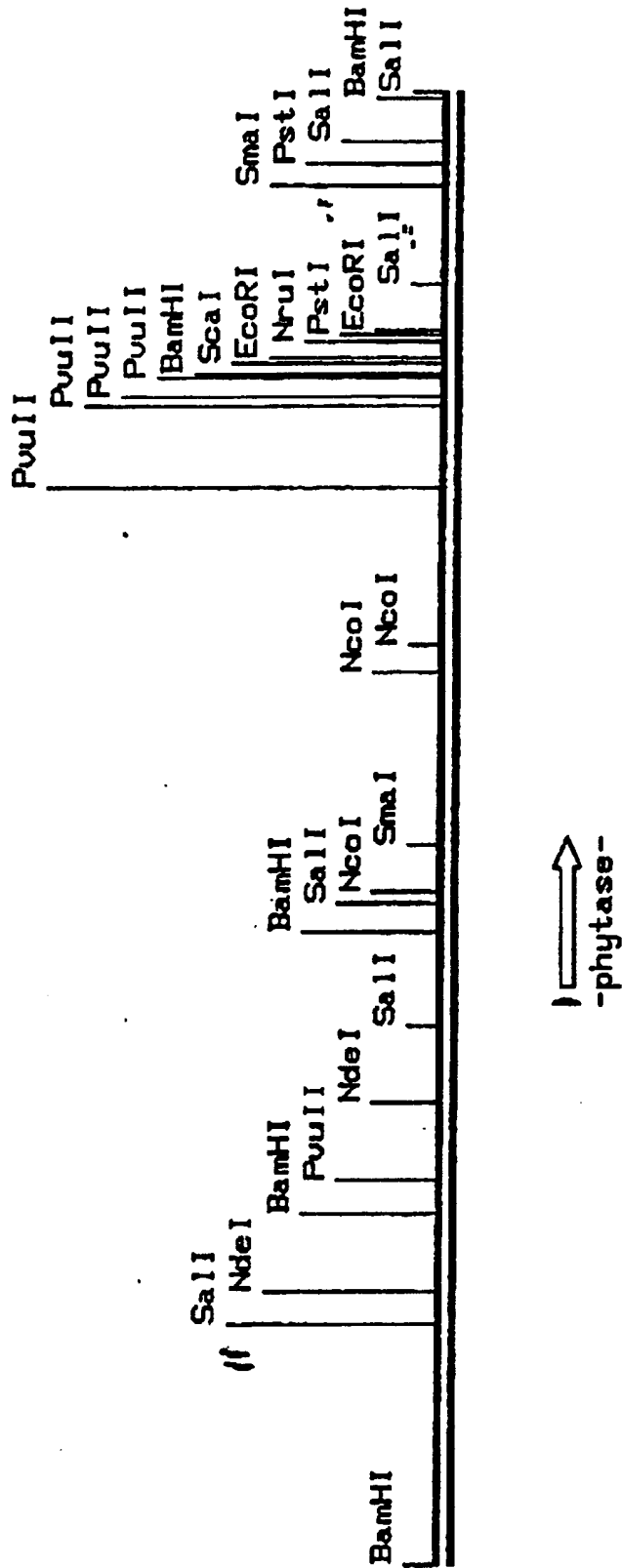


Figure 7

Figure 8

1 ATGGGCGTCTCTGCTGTTTCTACTTCCTTTGTATCTCCTGTCTGGAGTCACCTCCGGACTG  
-23 M Q V S A V L L P L Y L L S G V T S G L

61 GCAGTCCCCGCTCGAGAAATCAATCCAGTTGCGATACGGTCGATCAGGGGTATCAATGC  
A V P A S R N Q S S C D T V D Q G Y Q C  
-1 +1

121 TTCTCCGAGACTTCGCATCTTTGGGGTCAATACGCACCGTTCTTCTCTCTGGCAAACGAA  
18 F S E T S H L W G Q Y A P P F S L A N E

181 TCGGTCATCTCCCTGAGGTGCCCGCCGGATGCAGAGTCACTTTCCGTCAGGTCTCTCC  
38 S V I S P E V P A G C R V T F A Q V L S

241 CGTCATGAGCGCGGTATCCGACCGACTCCAAGGGCAAGAAATACTCCGCTCTCATTGAG  
58 R H Q A H Y P T D S K G K K Y S A L I E

301 GAGATCCAGCAGAACGCGACCACTTTGACGGAAAATATGCCCTTCTGAAGACATACAAC  
78 E I Q Q N A T T F D G K Y A P L K T Y N

361 TACAGCTTGGGTCCAGATGACCTGACTCCCTTCGGAGAACAGGAGCTAGTCAACTCCGGC  
98 Y S L G A D D L T P F G E Q E L V N S G

421 ATCAAGTTCTACCAACGGTACGAATCGCTCACAAGGAACATCGTTCCATTTCATCCGATCC  
118 I K F Y Q R Y E S L T R N I V P F I R S

481 TCTGGCTCCAGCCCGGTGATCCCTCCGGCAAGAAATTCATCGAGGGCTTCCAGACACC  
138 S G S S R V I A S G K K F I E G F Q S T

541 AAGCTGAAGGATCCTCGTGCCAGCCCGGCAATCGTCCGCCAAGATCGACGTGGTCATT  
158 K L K D P R A Q P G Q S S P K I D V V I

601 TCCGAGGCCAGCTCATCCAACAACACTCTCGACCCAGGCACCTGCACTGTCTTCGAAGAC  
178 S E A S S S N N T L D P G T C T V F E D

661 AGCGAATTGGCCGATACCGTCGAAGCCAATTCACCGCCACGTTGGTCCCCTCCATTGGT  
198 S E L A D ~~NT~~ V E A N F T A T F V P S I R

721 CAACGTCTGGAGAACGACCTGTCCGGTGTGACTCTCACAGACACAGAAGTGACCTACCTC  
218 Q R L E N D L S G V T L T D T E V T Y L

781 ATGGACATGTGCTCCTTCGACACCATCTCCACCAGCACCGTCGACACCAAGCTGTCCCCC  
238 M D N C S F D T I S T S T V D T K L S P

Figure 8 (continued 1)

841 TTCTGTGACCTGTTACCCCATGACGAATGGATCAACTACGACTACCTCCAGTCCTTGAAA  
 258 F C D L F T H D E W I N Y D Y L Q S L K

901 AAGTATTACGGCCATGGTGCAGGTAACCCGCTCGGCCCGACCCAGGGCGTCGGCTACGCT  
 278 K Y Y G H G A G N P L G P T Q G V G Y A

961 AACGAGCTCATCGCCCGTCTGACCCACTCGCCTGTCCACGATGACACCAGTTCCAACCAC  
 298 N E L I A R L T H S P V H D D T S S N H

1021 ACTTTGGA CTGAGCCCGGTACCTTTCCGCTCAACTCTACTCTCTACGCGGACTTTTCG  
 318 T L D S S P A T F P L N S T L Y A D F S

1081 CATGACAACGGCATCATCTCGATTCTCTTTGCTTTAGGTCTGTACAACGGCACTAAGCCG  
 338 H D N G I I S I L F A L G L Y N G T K P

1141 CTATCTACCAGCACC GTGGAGAATATCACCCAGACAGATGGATTCTCGTCTGCTTGACG  
 358 L S T T T V E N I T Q T D G F S S A W T

1201 GTTCCGTTTGCTTCGGCTTTGTACCTCGAGATGATGCAGTGTACGGCGAGCAGGAGCCG  
 378 V P F A S R L Y V E M M Q C Q A E Q E P

1261 CTGGTCCGTGTCTTGGTTAATGATCGCGTTGTCCCGCTGCATGGGTGTCCGGTTGATGCT  
 398 L V R V L V N D R V V P L H G C P V D A

1321 TTGGGGAGATGTACCCGGGATAGCTTTGTGAGCGGGTTGAGCTTTGCTAGATCTGGGGGT  
 418 L G R C T R D S F V R G L S F A R S G G

1381 GATTGGCCGGAGTGT TTTGCTTAG  
 438 D W A E C F A \*

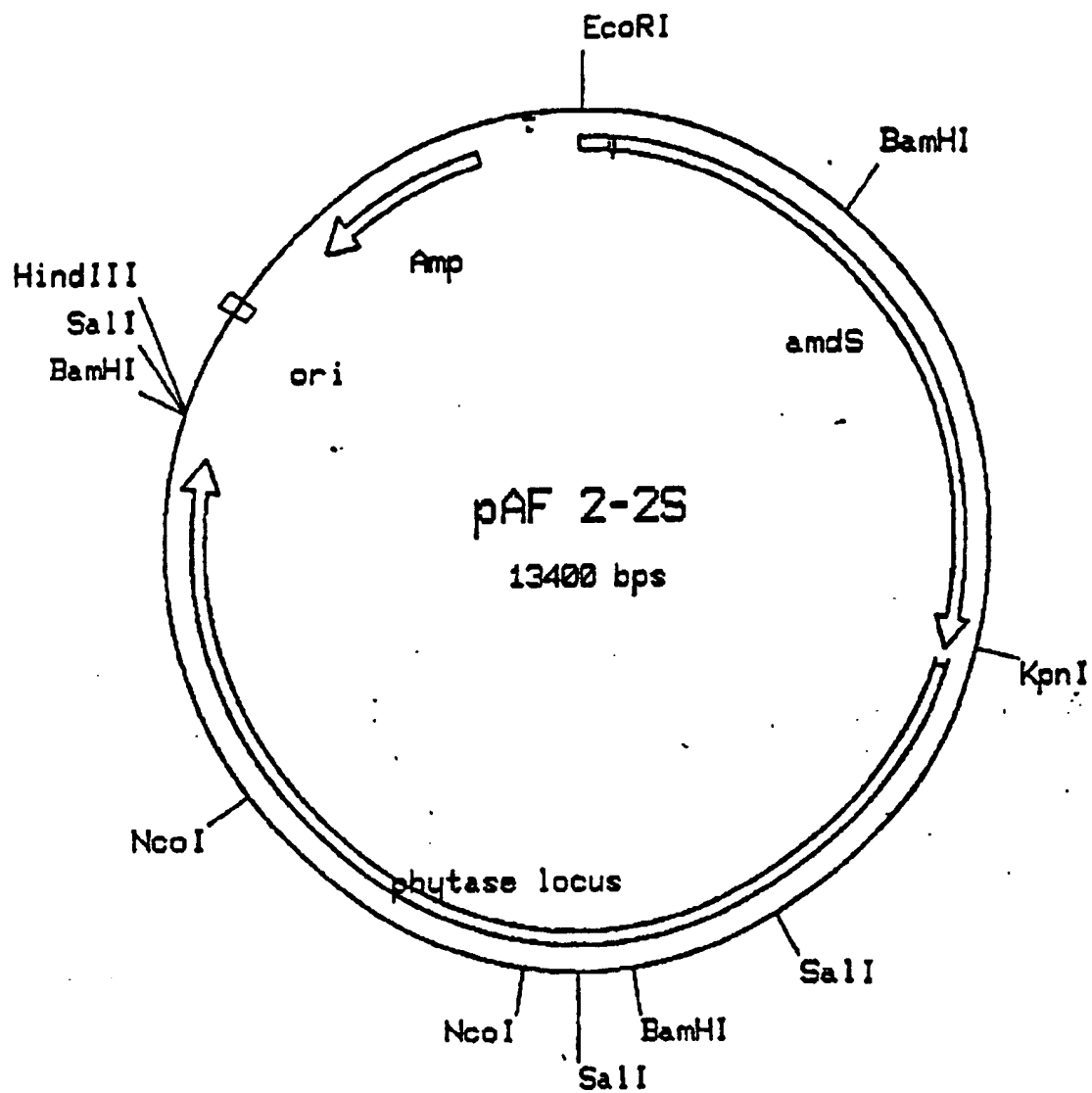


Figure 9

A

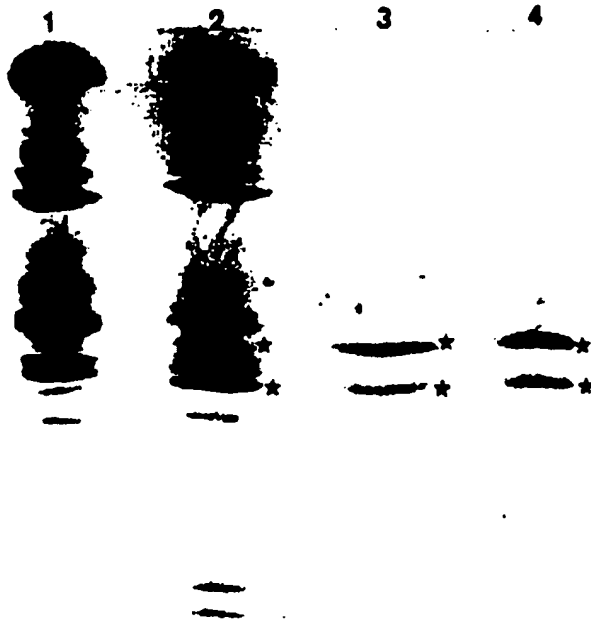


Figure 10A

B

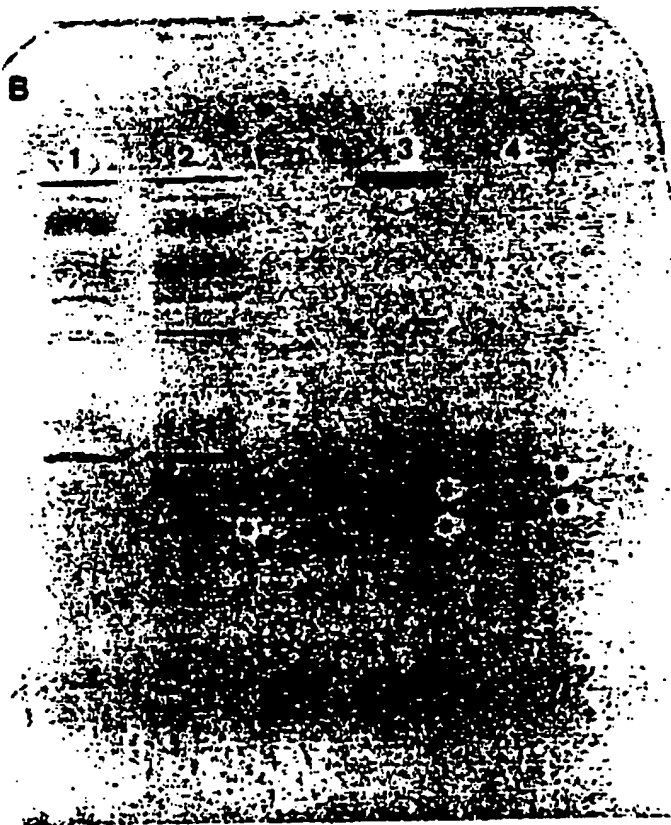


Figure 10B

A

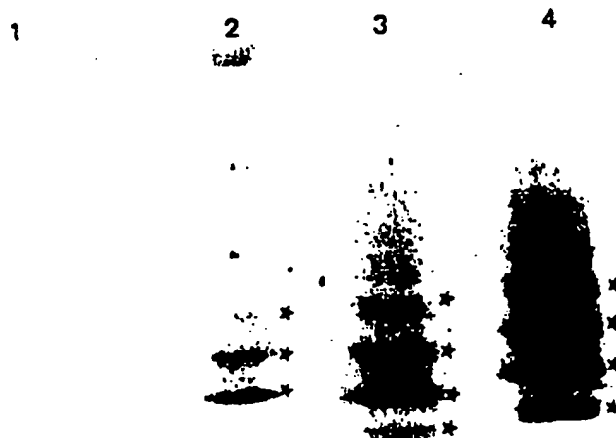


Figure 11A

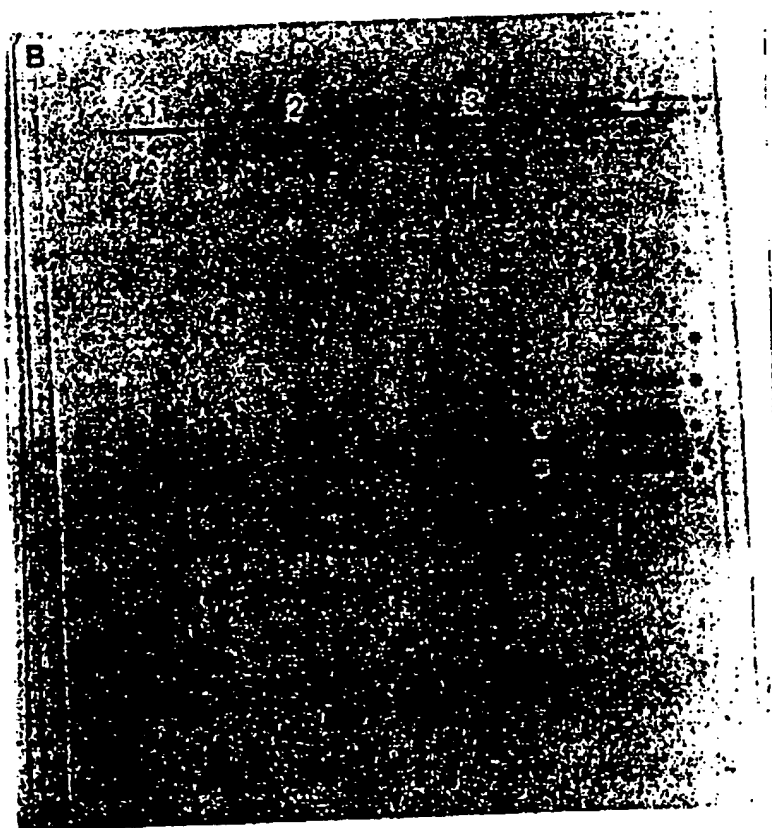


Figure 11B



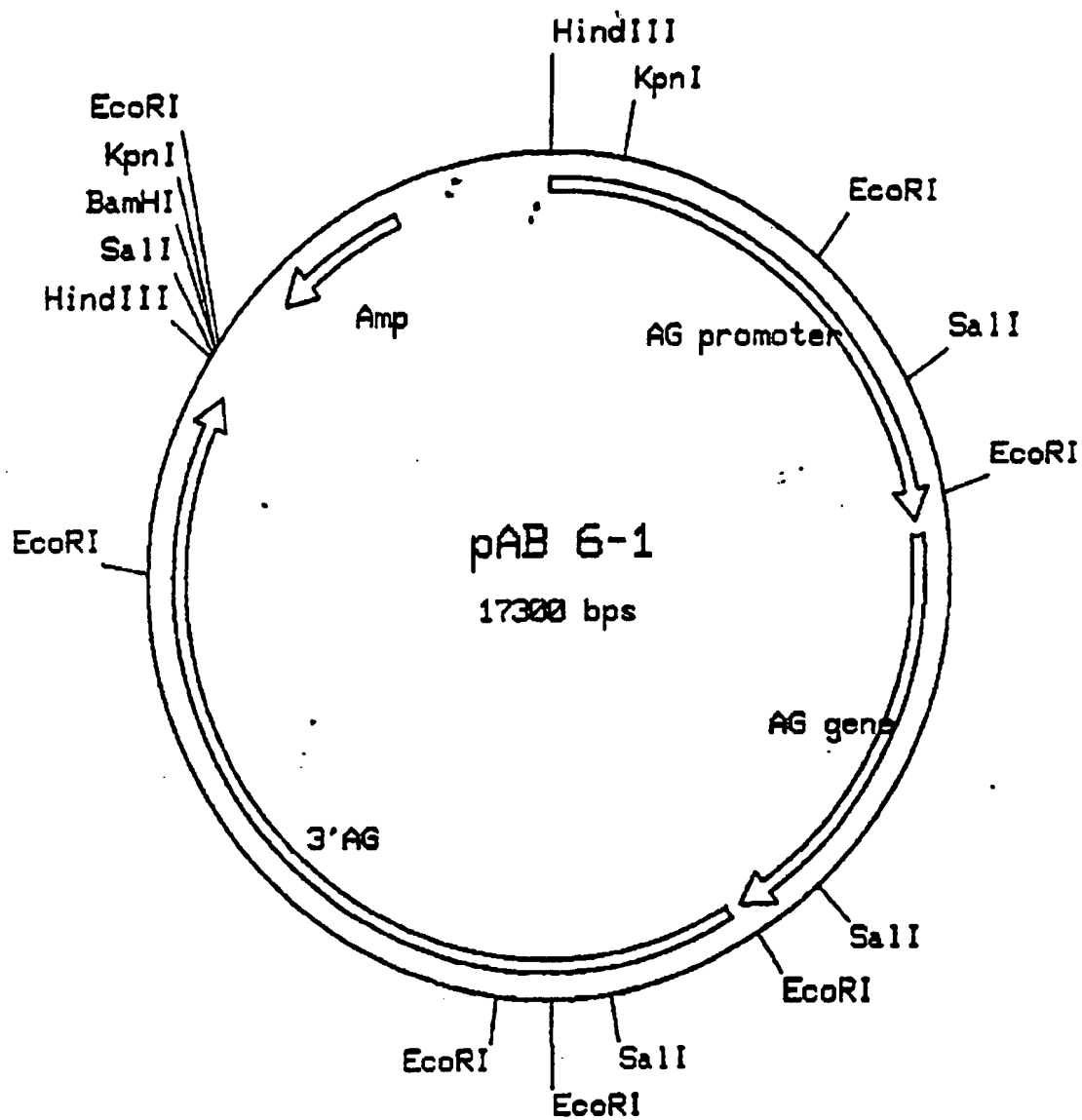


Figure 12

## AG/PHYTASE GENE FUSIONS BY PCR

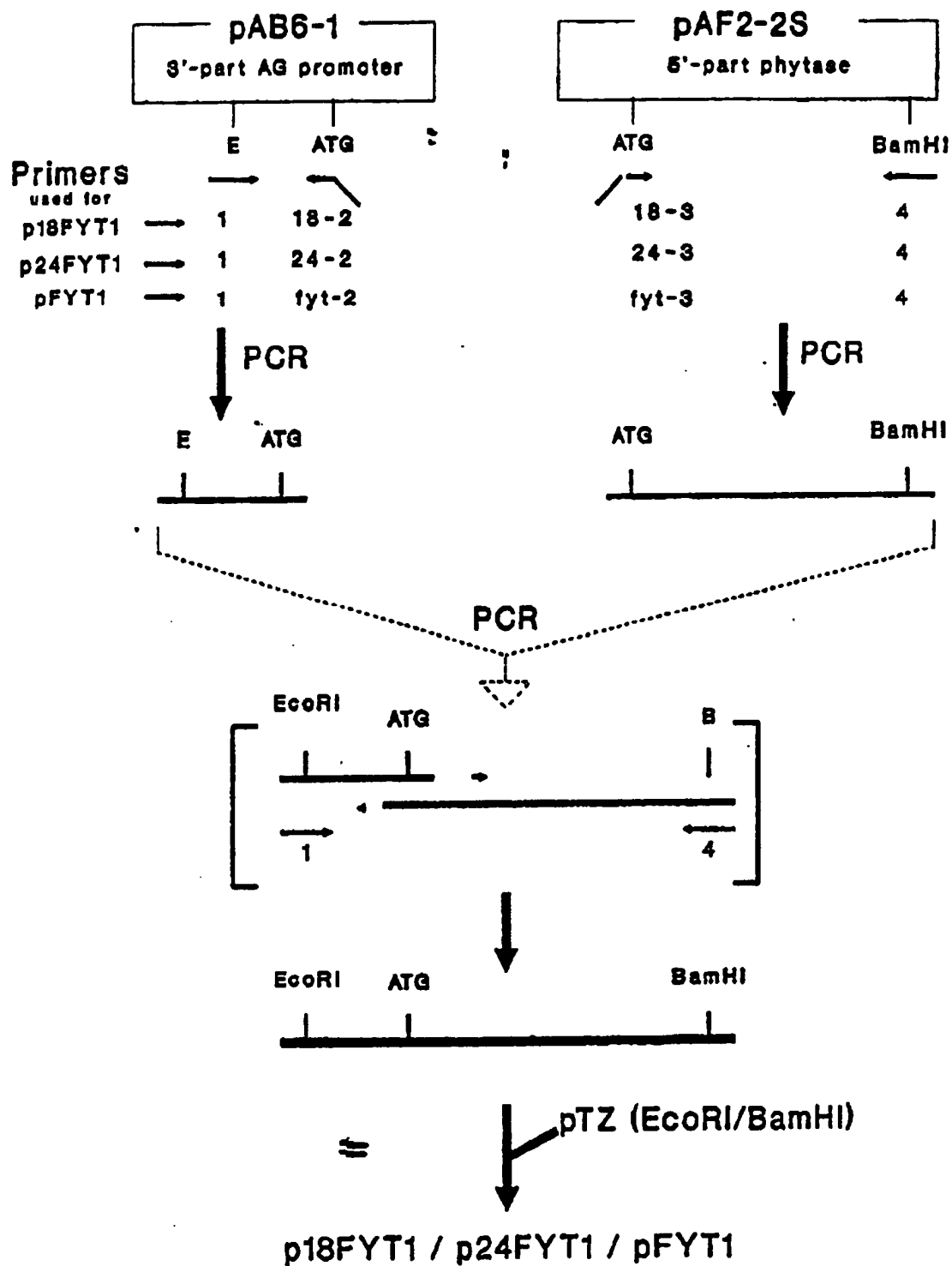


Figure 13

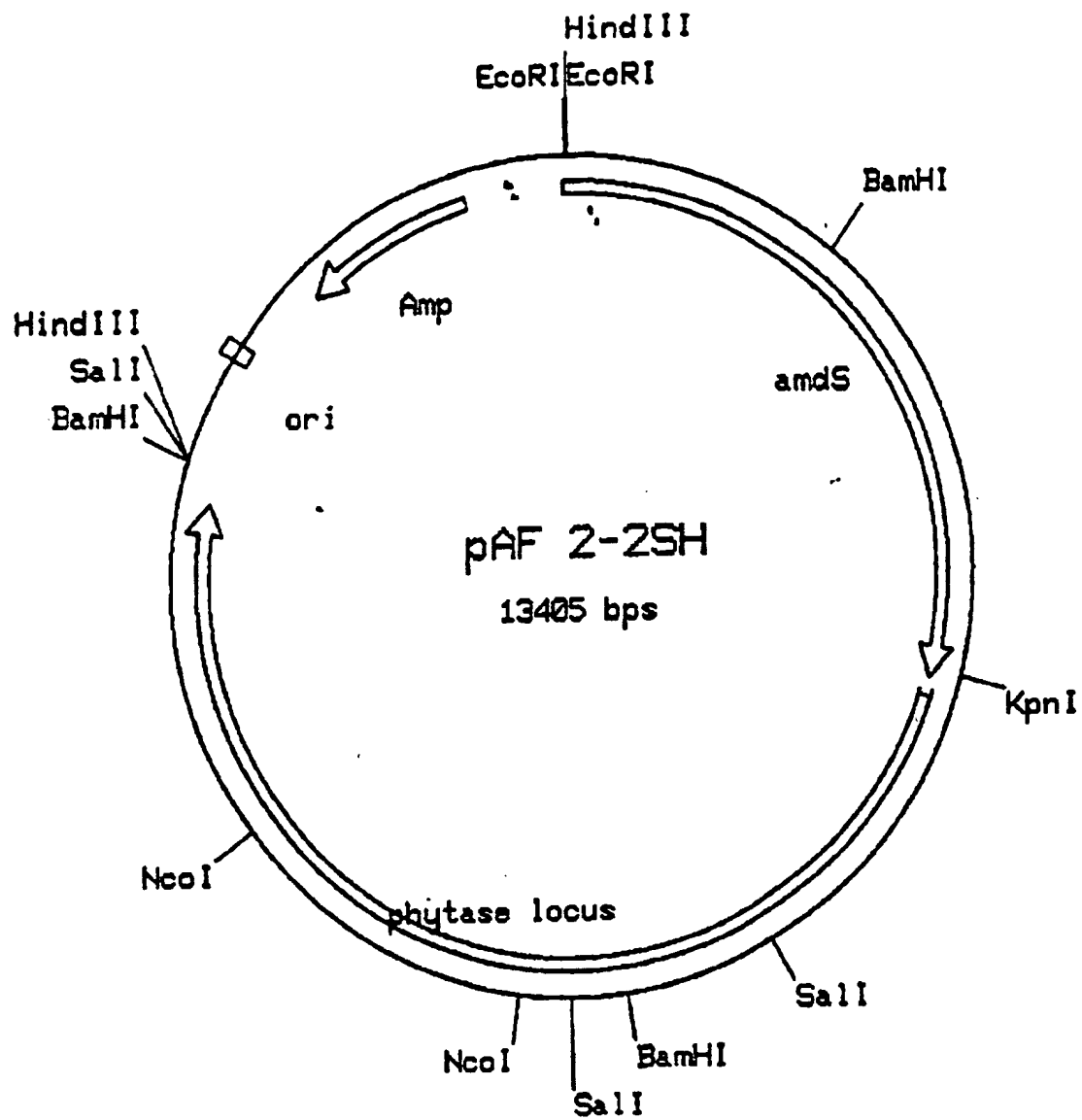


Figure 14

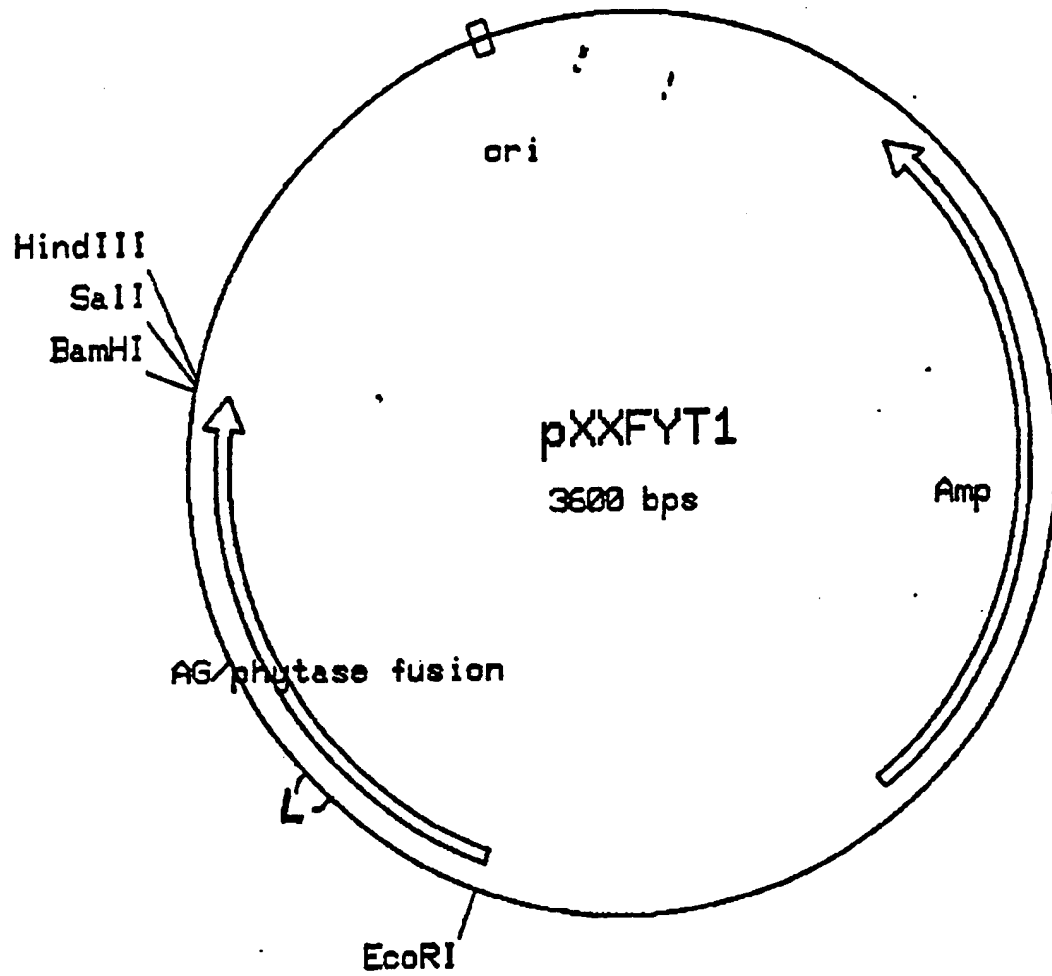


Figure 15a

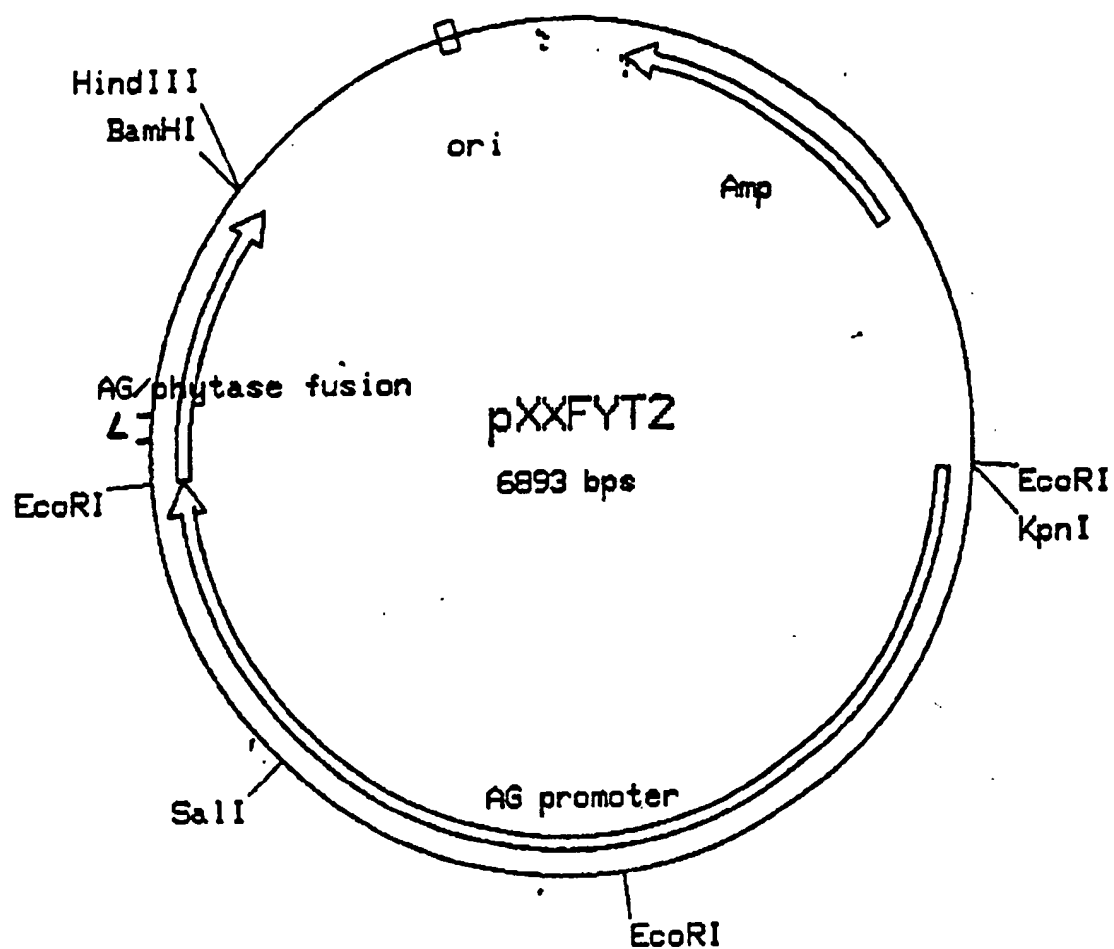


Figure 15b

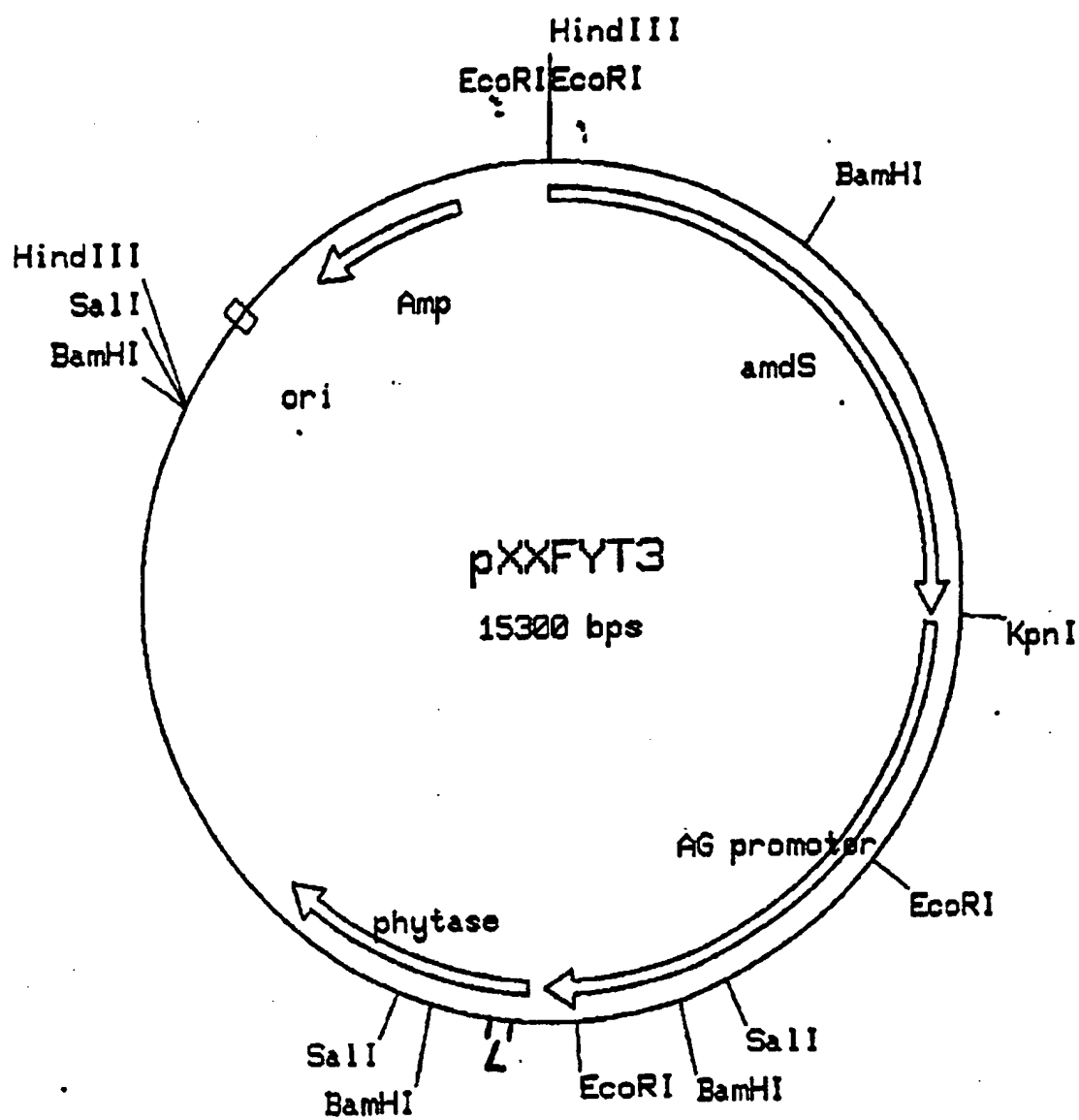


Figure 15c

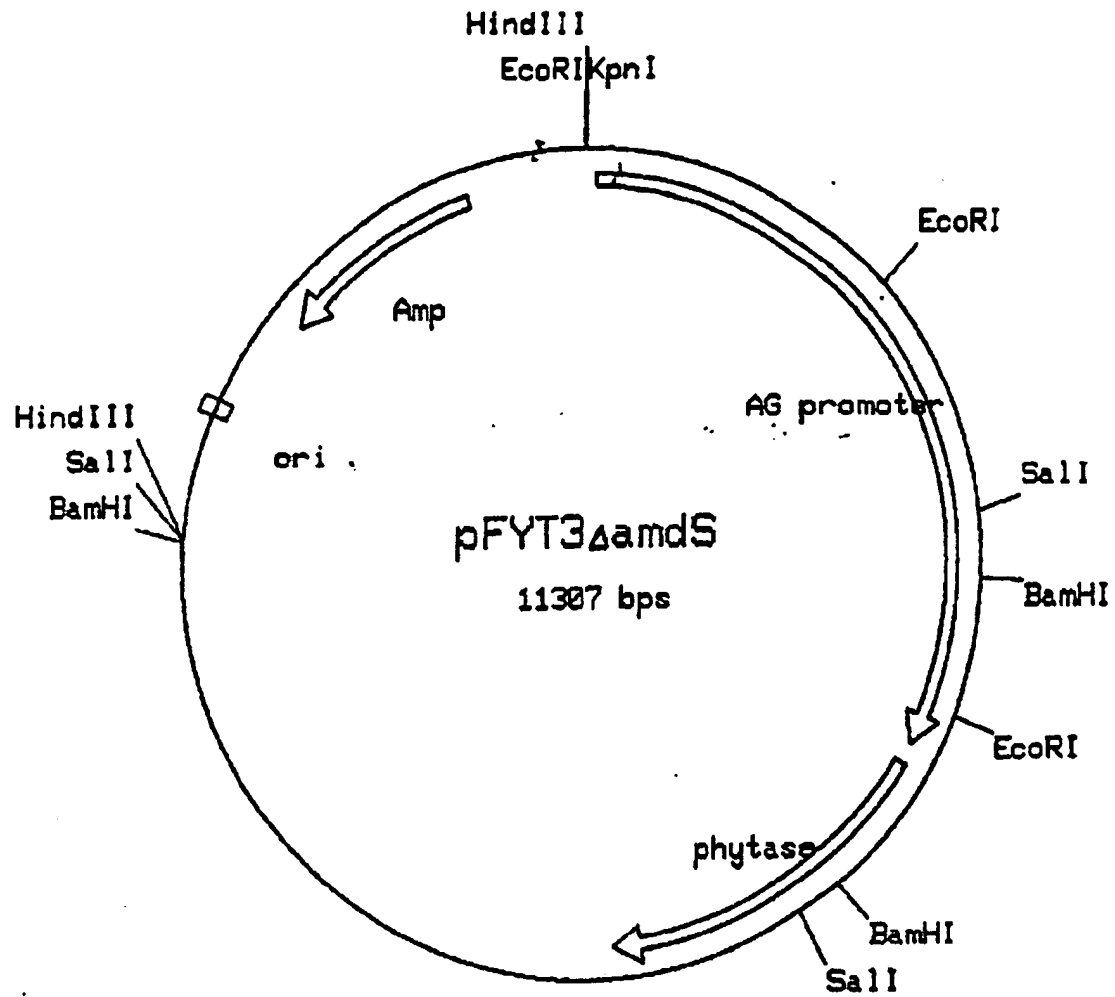


Figure 16

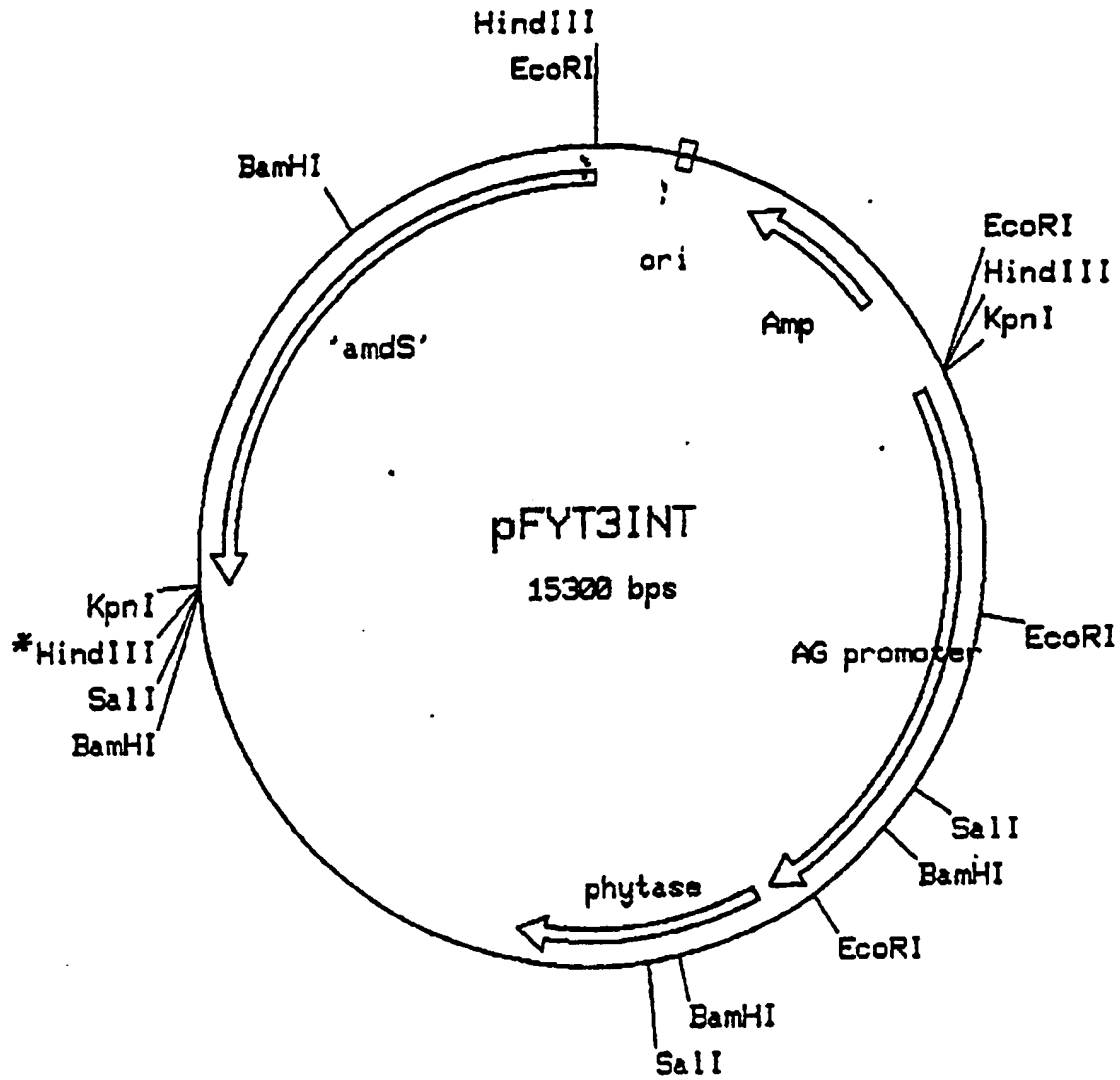


Figure 17



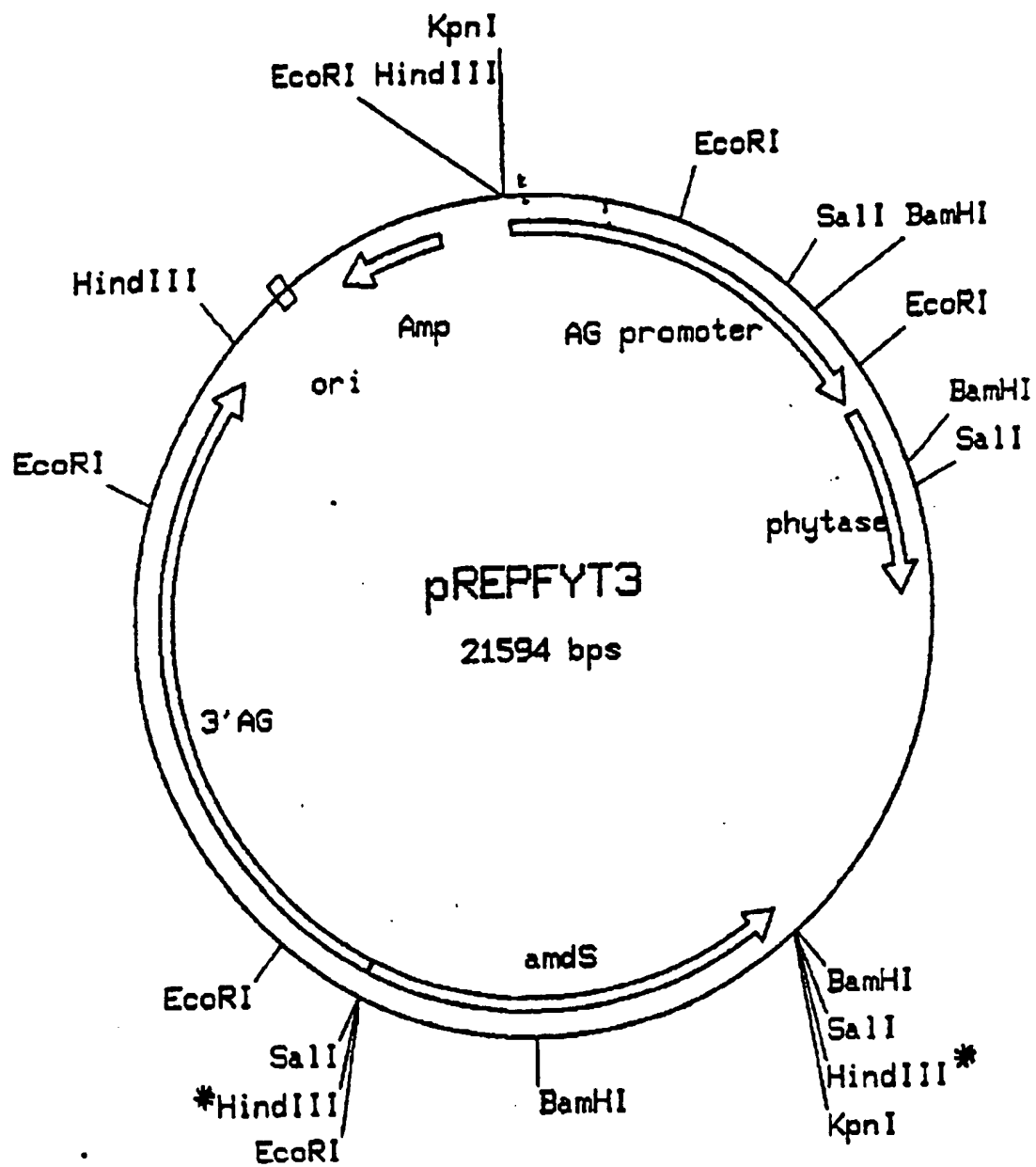


Figure 18

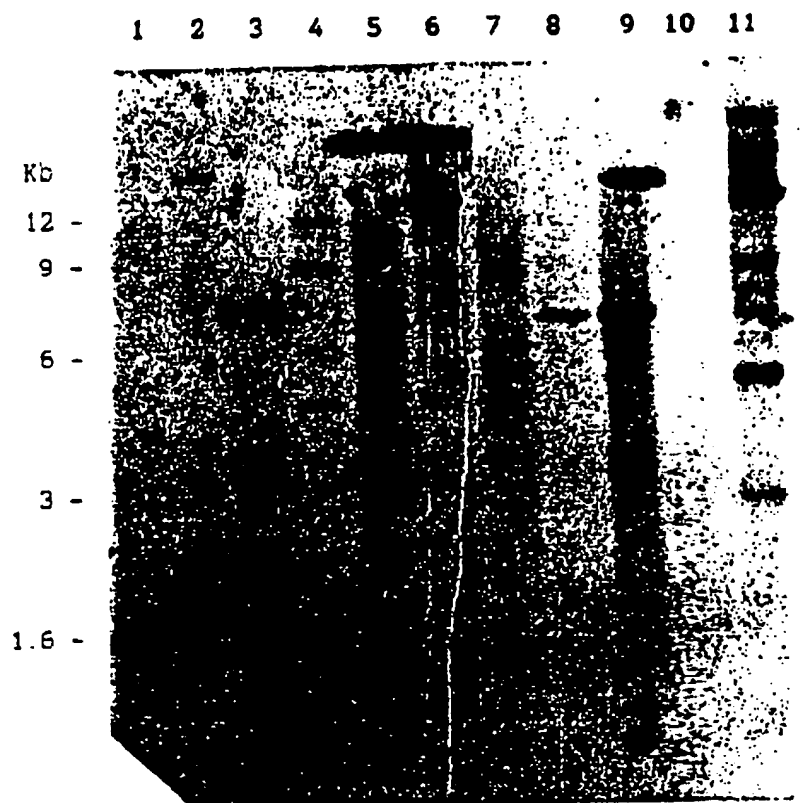


FIGURE 19 A

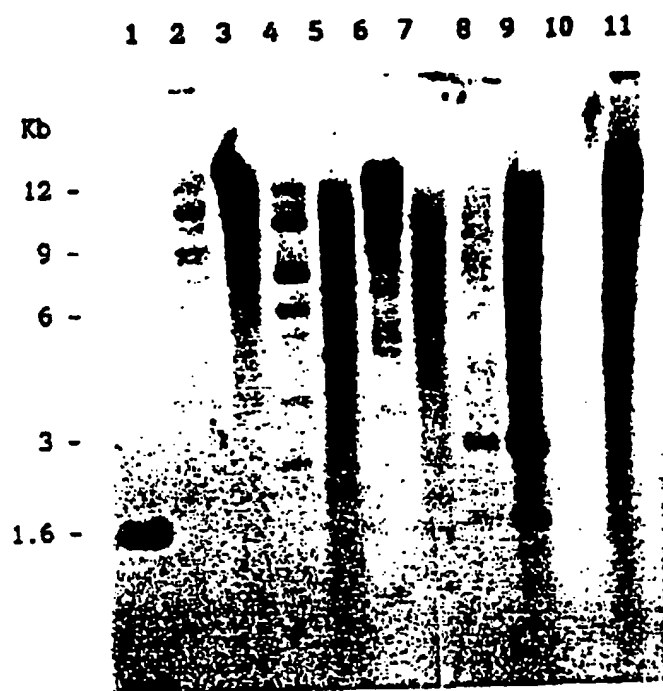


FIGURE 19 B



European  
Patent Office

## EUROPEAN SEARCH REPORT

Application Number

EP 90 20 2565

DOCUMENTS CONSIDERED TO BE RELEVANT			
Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int. Cl.5)
D,Y	PREPARATIVE BIOCHEMISTRY, vol. 18, no. 4, 1988, pages 458-471; A.H.J. ULLAH: "Aspergillus ficuum phytase: partial primary structure, substrate selectivity, and kinetic characterization" * Whole document *	1-27	C 12 N 15/55 C 12 N 9/18 A 23 K 1/165 C 12 N 1/15 C 12 N 1/19
Y	EP-A-0 287 152 (GIST-BROCADES) * Column 4, line 44 - column 5, line 34; claim 17 *	1-27	
D,A	NATURE, vol. 318, 14th November 1985, pages 191-194; S.M. SAMSON et al.: "Isolation, sequence determination and expression in Escherichia coli of the isopenicillin N synthetase gene from Cephalosporium acremonium" * Whole document *	1	
D,A	US-A-3 297 548 (J.H. WARE et al.) * Whole document *	28-31	
			TECHNICAL FIELDS SEARCHED (Int. Cl.5)
			C 12 N
The present search report has been drawn up for all claims			
Place of search		Date of completion of search	Examiner
The Hague		14 January 91	DELANGHE L.L.M.
<b>CATEGORY OF CITED DOCUMENTS</b>			
X: particularly relevant if taken alone		E: earlier patent document, but published on, or after the filing date	
Y: particularly relevant if combined with another document of the same category		D: document cited in the application	
A: technological background		L: document cited for other reasons	
O: non-written disclosure		S: member of the same patent family, corresponding document	
P: intermediate document			
T: theory or principle underlying the invention			

## Plant Gene Register

## Starch Branching Enzyme II from Maize Endosperm

Dane K. Fisher, Charles D. Boyer\*, and L. Curtis Hannah

Department of Horticulture, The Pennsylvania State University, University Park, Pennsylvania 16802 (D.F., C.D.B.); and Department of Vegetable Crops, The University of Florida, Gainesville, Florida 32611 (L.C.H.)

ADP-Glc pyrophosphorylase (EC 2.7.7.27), starch synthases (EC 2.4.1.21), and SBEs (EC 2.4.1.18) are the key enzymes in the pathway of plant starch biosynthesis. ADP-Glc pyrophosphorylase and starch synthase catalyze the formation of ADP-Glc, the substrate for starch synthesis, and the polymerization of Glc into the amylose and amylopectin fractions of the starch granule, respectively. SBEs catalyze branch point formation by the cleavage and reattachment of  $\alpha$ -1,4-linked Glc chains to  $\alpha$ -1,6 branch points in the growing starch molecule (Borovsky et al., 1979; Boyer, 1985). Branching enzymes are proposed to interact with starch synthases in formation of amylopectin (Boyer and Preiss, 1981).

Multiple forms of SBE have been identified in maize (*Zea mays* L.) endosperm (Boyer and Preiss, 1978, 1981). Three forms of SBE, I, IIa, and IIb, from developing maize endosperm have been characterized by their level of branching activity on amylose and amylopectin and by kinetic and immunological properties (Boyer and Preiss, 1978; Fisher and Boyer, 1983; Singh and Preiss, 1985). Distinct differences have been shown between SBE I and IIa or IIb by the above parameters, but only small differences exist between SBE IIa and IIb. Genetic evidence suggests that IIa and IIb are products of separate genes (Boyer and Preiss, 1981; Hedman and Boyer, 1982). The precise role of each isoform in starch formation has yet to be determined. To further understand these enzymes, efforts to clone the genes have been undertaken. We report here the cloning of a SBE II cDNA from maize (Table I). Three  $\lambda$ gt10 cDNA libraries were constructed from endosperm poly(A)<sup>+</sup> RNA 14, 22, and 29 DAP. A heterologous nucleic acid probe, clone pJSBE5, the cDNA for pea SBE I, was used to screen the 14-DAP library (Bhattacharyya et al., 1990). After purifying and subcloning into plasmid pBluescript II SK<sup>-</sup> (Stratagene), a full-length cDNA of 2725 bp was isolated.

Northern blots of total maize RNA isolated from endosperm tissue 12 DAP and probed with the cloned maize cDNA revealed a single transcript of approximately 2.7 kb. Deduced amino acid sequence was compared with the pea SBE I (Bhattacharyya et al., 1990), maize SBE I (Baba et al., 1991), and rice SBE I (Nakamura et al., 1992) translated cDNA sequences using Intelligenetics software. Levels of residue identity were 71, 52, and 52%, respectively. From these results, we conclude that we have cloned a second

**Table I.** Characteristics of SBE II cDNA from maize endosperm

Organism:
<i>Zea mays</i> L. (W64A $\times$ 182E).
Gene Product:
SBE (1,4- $\alpha$ -D-glucan 6- $\alpha$ -D-(1,4- $\alpha$ -D-glucanotransferase); EC 2.4.1.18); starch biosynthesis.
Clone Type; Designation:
cDNA, full-length; $\lambda$ 29-III-1; pMA11(pBluescript).
Source:
cDNA libraries in $\lambda$ gt10 constructed from maize endosperm poly(A) <sup>+</sup> mRNA isolated 14, 22, and 29 DAP.
Techniques:
Libraries screened with pea ( <i>Pisum sativum</i> L.) SBE I clone pJSBE5 (Bhattacharyya et al., 1990); three overlapping clones subcloned into pBluescript II SK <sup>-</sup> ; double-stranded dideoxynucleotide sequencing of overlapping clones using various subclones and synthetic oligonucleotide primers; second screening with 5' and 3' ends of partial clones to obtain full-length clone; both strands of the full-length clone were sequenced ( $\lambda$ 29-III-1; pMA11).
Method of Identification:
Sequence homology to other SBE clones; deduced amino acid (residues 58-65) identity to purified maize SBE IIb mature protein N-terminal sequence.
Structural Features of Protein:
Open reading frame of 798 amino acids; calculated <i>M<sub>r</sub></i> of mature protein of 84,772; putative 53-amino acid transit peptide N terminal to mature protein sequence.
Subcellular location:
Amyloplast.

isoform of SBE from maize endosperm. This conclusion is supported by the N-terminal sequence of purified maize SBE IIb protein, which matches the cDNA predicted amino acid sequence at residues 58 to 65. The additional amino acid residues making up the N-terminal end of the deduced sequence are thought to encode a transit peptide (53 amino acids) for routing of the protein to the amyloplast. The deduced molecular mass of the mature protein from this sequence data is 84,772 D. This is slightly larger than size estimates of 80,000 D based upon SDS-PAGE analysis of

\* Corresponding author; fax 1-814-863-6139.

purified SBE IIa and IIb protein (Boyer and Preiss, 1978; Singh and Preiss, 1985).

#### ACKNOWLEDGMENTS

We thank Dr. Cathie Martin (John Innes Institute, Norwich, UK) for providing pea clone pJSBE5 and Dr. Lynn Ingram (University of Florida) for providing the endosperm cDNA libraries.

Received December 22, 1992; accepted January 9, 1993.

Copyright Clearance Center: 0032-0889/93/102/1045/02.

The GenBank/EMBL accession number for the sequence reported in this article is L08065.

#### LITERATURE CITED

- Baba T, Kimura K, Mizuno K, Etoh H, Ishida Y, Shida O, Arai Y (1991) Sequence conservation of the catalytic regions of amylolytic enzymes in maize branching enzyme-I. *Biochem Biophys Res Commun* 181: 87-94
- Bhattacharyya MK, Smith AM, Ellis THN, Hedley C, Martin C (1990) The wrinkled-seed character of pea described by Mendel is caused by a transposon-like insertion in a gene encoding starch-branching enzyme. *Cell* 60: 115-122
- Borovsky D, Smith E, Whelan W, French D, Kikumoto S (1979) The mechanism of Q-enzyme action and its influence on the structure of amylopectin. *Arch Biochem Biophys* 198: 627-631
- Boyer CD (1985) Synthesis and breakdown of starch. In C Neyra, ed, *Biochemical Basis of Plant Breeding*, Vol I. CRC Press, Boca Raton, FL, pp 133-153
- Boyer CD, Preiss J (1978) Multiple forms of (1,4)-alpha-D-glucan, (1,4)-alpha-D-glucan-6-glycosyl transferase from developing *Zea mays* L. kernels. *Carbohydr Res* 61: 321-334
- Boyer CD, Preiss J (1981) Evidence for independent genetic control of the multiple forms of maize endosperm branching enzymes and starch synthases. *Plant Physiol* 67: 1141-1145
- Fisher MB, Boyer CD (1983) Immunological characterization of maize starch branching enzymes. *Plant Physiol* 72: 813-816
- Hedman KD, Boyer CD (1982) Gene dosage at the *amylose-extender* locus of maize: effects on the levels of starch branching enzymes. *Biochem Genet* 20: 483-491
- Nakamura Y, Yamanouchi H (1992) Nucleotide sequence of a cDNA encoding starch-branching enzyme, or Q-enzyme I, from rice endosperm. *Plant Physiol* 99: 1265-1266
- Singh BK, Preiss J (1985) Starch branching enzymes from maize. Immunological characterization using polyclonal and monoclonal antibodies. *Plant Physiol* 79: 34-40

Structure of the sucrose synthase gene on chromosome 9 of *Zea mays* L.

W.Werr, W.-B.Frommer, C.Maas and P.Starlinger

Institut für Genetik der Universität Köln, 5000 Köln 41, FRG

Communicated by P.Starlinger

The structure of the shrunken gene of *Zea mays* encoding sucrose synthase (EC 2.4.1.13) was determined by (i) sequencing the transcription unit and ~1.2 kb of 5'-upstream sequences from a genomic clone, (ii) by sequencing a nearly full length cDNA clone and (iii) by determining the transcription start site by a combination of primer extension experiments with synthetic oligodeoxynucleotide primers and S1 mapping. The sucrose synthase gene is 5.4 kb long, of which 2746 bp are found in the mature mRNA. The gene is interrupted by 15 introns. The first two introns are ~1 kb and ~0.5 kb in length, respectively, while the other introns are much smaller. A TATA box is located 30 bp upstream from the transcription start site. Approximately 610 bp upstream of the transcription start site a direct repeat of 16 nucleotides, separated by a 4-fold repetition of the sequence GGTGG is detected. The 16-bp sequence has similarities to a sequence repeat found between two promoters of a maize zein gene also expressed in the endosperm tissue. The transposable element *Ds* in the mutant *sh-m5933* and *sh-m6233* alleles is inserted in the seventh and first intron, respectively. The genomic and cDNA clones were obtained from different maize lines. This allows the determination of polymorphic sites which are frequent in 3rd codon position and absent in 1st and 2nd codon positions. In addition, the 3'-untranslated sequence shows two duplications that may have arisen by the insertion and subsequent excision of transposable elements. **Key words:** sucrose synthase/transcription signals/sequence polymorphism/transposon footprints

## Introduction

The *Shrunken* (*Sh*) locus on chromosome 9 of *Zea mays* encodes the enzyme sucrose synthase (EC 2.4.1.13) which catalyzes the cleavage reaction of sucrose to UDP-glucose and fructose in both directions (Chourey and Nelson, 1976). The enzyme is involved in starch metabolism of the developing endosperm. In maize strains homozygous for recessive *sh* mutations, sucrose synthase activity is decreased to 2-6% and starch content in the mature kernels is decreased to 60%. This causes the *shrunken* phenotype (Chourey and Nelson, 1976; Chourey, 1981). In wild-type, the protein encoded by the *Sh* locus amounts up to 3% of the total protein. The active form of the enzyme consists of four identical subunits of ~88 kd (Tsai, 1974). It is not known whether the high residual starch content in the mutants indicates that the enzyme is present in large excess in the wild-type or whether alternative pathways of starch biosynthesis are used.

A residual sucrose synthase (B) activity is detected in *sh* mutants (Chourey, 1981; Chourey and Nelson, 1976), including a deletion of the *Sh* gene (Burr and Burr, 1981; Chaleff *et al.*, 1981; Döring *et al.*, 1981). This indicates the presence of a

second gene. In support of this idea, McCormick *et al.* (1982) detected a mRNA species and DNA fragments hybridizing weakly to a cDNA clone of the *Sh* gene in RNA and DNA isolated from maize strains homozygous for *Sh* deletions. The *in vitro* translation product of hybrid-selected mRNA transcribed from the sucrose synthase (B) gene has a similar, slightly faster electrophoretic mobility than the *in vitro* synthesized sucrose synthase (A) subunit encoded by the *Sh* locus and is precipitated by antiserum against sucrose synthase (A) protein (McCormick *et al.*, 1982).

The two genes seem to be regulated differently. In wild-type strains sucrose synthase A activity increases 40-fold during kernel development in the endosperm tissue, starting at day 5-8 after pollination. The enzyme activity reaches a maximum 40 days after fertilization and drops afterwards (Chourey, 1981). In strains in which the *Sh* gene is deleted the residual sucrose synthase B activity stays at a low level. Independent of the genotype (*Sh* or *sh*), only sucrose synthase B is detected in the embryo of the developing kernel (Chourey and Nelson, 1976).

A second reason for the interest in studying expression of sucrose synthase genes is the reaction catalyzed by the enzyme. Sucrose is the major transport form of assimilate on its way from the photosynthesizing leaves to the energy-consuming tissues, e.g., to the kernels. The observations from several plants are in agreement with the hypothesis that synthesis of sucrose is mainly catalyzed by sucrose-6-phosphate synthase (EC 2.4.1.14) and sucrose phosphatase (EC 3.1.3.24) and that sucrose synthase has its major role in sucrose breakdown for respiration or starch biosynthesis (Hawker, 1971; Downton and Hawker, 1973; Vieweg, 1974; Preiss and Levi, 1980). The enzyme therefore may be involved in the utilization of sucrose within different plant tissues, and may thus influence where the sucrose is stored or broken down, respectively.

As a step towards the study of gene regulation during plant development, we present here the exon-intron structure of the *Sh* gene. In addition, we discuss some evolutionary aspects deduced from the comparison of different alleles of the *Shrunken* gene and DNA sequences upstream of the transcription unit which might be important for regulation.

## Results

DNA sequence studies of the *Shrunken* gene

For our sequence studies we used one genomic and two cDNA clones. A 600-bp 3'-terminal cDNA clone and a 16.3-kb genomic clone have been described previously (Geiser *et al.*, 1980, 1982). A 2.572-bp cDNA clone (pWW110/1) was isolated from a cDNA library kindly provided by A.Gierl and Zs.Schwarz-Sommer, Max-Planck-Institut für Züchtungsforschung, Cologne. On the genomic clone, the whole transcription unit and a DNA segment of 1140 bp located in front of the transcription unit were sequenced from both strands. The small cDNA clone was also sequenced from both strands, while from the larger cDNA clone only one strand was sequenced. Comparison of the two clones yielded the structure shown in Figure 1, where the cDNA clone

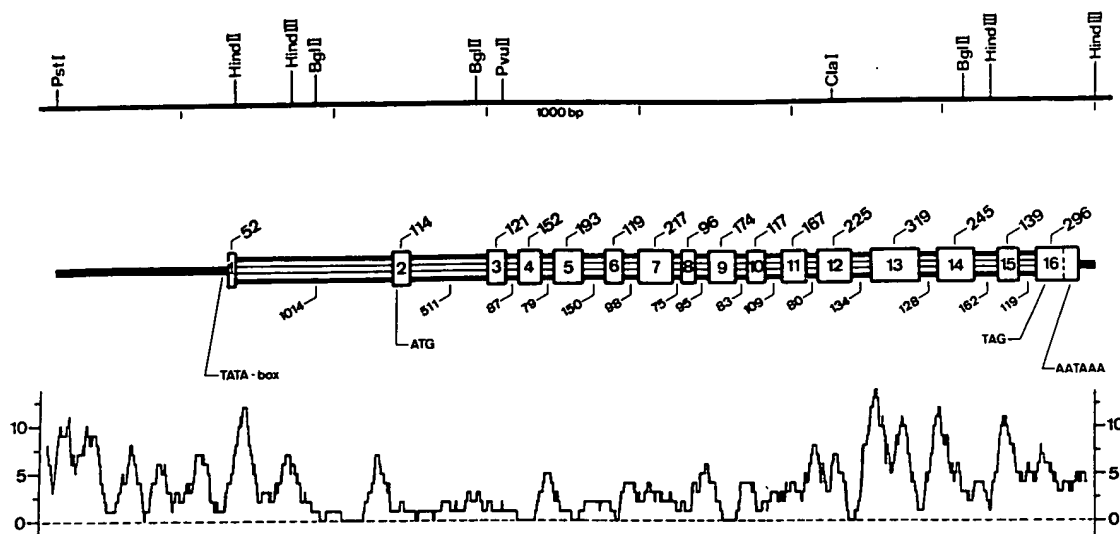


Fig. 1. Exon/intron structure of the *shrunken* gene. The exons are marked by the numbered boxes. The size of the exons is given above the drawing, intron sizes below. The total length of the drawing indicates the region which has been sequenced. A partial restriction map of the genomic DNA is shown in the upper part. The distribution of the dinucleotide CpG is given in absolute numbers for consecutive segments of 100 bp each.

(pWW110/1) ends within exon 3. The full sequence of both clones is entered in the EMBL DNA library and is available from the authors upon request.

#### Determination of the transcription start site

The large cDNA clone is not full length and terminates in exon 3. This was determined by S1 mapping experiments which were carried out in order to determine the approximate size and the order of the transcription unit and the size of the exons. For these experiments, the 7.1-kb *ClaI* fragment was labeled at the 5' terminus. This fragment has one end in the central part of the transcribed region, extends beyond the transcription start and contains all upstream sequences present in our genomic wild-type clone. Aliquots of the labeled fragments were cleaved with restriction endonucleases *Bam*HI, *Bgl*II, *Xba*I, *Hind*III, *Sph*I and *Bgl*II. The fragments located towards the 3' end of the gene were purified, hybridized to mRNA, digested with S1 endonuclease, separated on agarose gels and subjected to autoradiography. The result of such an experiment is shown in Figure 2. As described in the sketch in Figure 2b, the longest band in each lane represents the full-length protected fragment, while the shorter bands are formed by partial S1 cleavage opposite to digested intron loops. The successive shortening of the labeled fragments by restriction digestion should have no influence on the length of the longest bands, until the first exon has been (partially) removed from the DNA before hybridization. By this method, the first exon is located in the 700-bp interval between the *Xba*I and the *Hind*III site (Figure 1), while the second exon is located between the *Hind*III and *Sph*I sites. The third exon is thus located within the central 3.3-kb *Bgl*II fragment. It contains a *Pvu*II site which is also present in the cDNA clone pWW110/1. The position of the intron 2/exon 3 boundary was determined by a S1 mapping experiment with DNA fragments labeled at the 5' terminus of this *Pvu*II site. The resulting S1-resistant DNA fragment was 40 bp long, thus the *Pvu*II site is located 40 bp downstream of the intron/exon boundary (indicated in Figure 3c). The first three exons are thus spread over a distance of > 1.5 kb. As we had no cDNA clone in this region available, we decided to perform primer extension experiments with reverse transcriptase.

Poly(A)<sup>+</sup> RNA isolated from developing endosperm was

reverse transcribed in the presence of the synthetic oligodeoxynucleotide AAGCATTCCCTTGCCC as primer. The position of the primer sequence within exon 3 is indicated in Figure 3c. Reverse transcription was done on poly(A)<sup>+</sup> RNA isolated either from maize strains carrying the *Sh* allele, or as a control from maize carrying the deletion *sh bz-m4*. Reverse transcripts were electrophoresed on 6% denaturing polyacrylamide gels. As shown in Figure 3, only the wild-type RNA gives rise to a prominent double band at position 191/192, while many faint bands are also present in the control. Most likely, this strong band is a reverse transcript extending to the 5' end of the mRNA. The size of exon 2 was estimated accurately from S1 mapping experiments to be 115 bp. In these experiments S1-digested RNA/DNA hybrids were denatured and electrophoresed on polyacrylamide gels containing 7 M urea. After transfer to nitrocellulose and hybridization to radioactively labeled DNA fragments, the length of the individual exons could be determined precisely.

If the known size of exon 2 and the residual bases of exon 3 towards the 5' end of the synthetic oligonucleotides are added and subtracted from the length of the reverse transcript (191/192 bases), there remain 50 or 51 nucleotides for the size of exon 1, which is in agreement with the size estimated from S1 mapping experiments (Figure 2).

To determine the position of exons 1 and 2, the primer extension experiment was repeated in the presence of dideoxynucleotides to determine the sequence of the reverse transcript. The autoradiographs obtained from this experiment showed a high background most likely due to unspecific priming of reverse transcriptase as already visible in the *sh bz-m4* lane of Figure 3a. Therefore only parts of the sequence could be unambiguously determined. In general, the sequence was less clear at the higher transcript sizes.

Comparing this limited cDNA sequence information with the known genomic DNA sequence, we could recognize ~70% of exon 2 sequence. The sequence information obtained at the boundary between exons 1 and 2 is shown in Figure 3c. The unambiguously determined nucleotides extending further than exon 2 were screened in a computer search against all DNA sequences which were available upstream of exon 2. Even with one mismatch allowed, homology was found only once as indi-



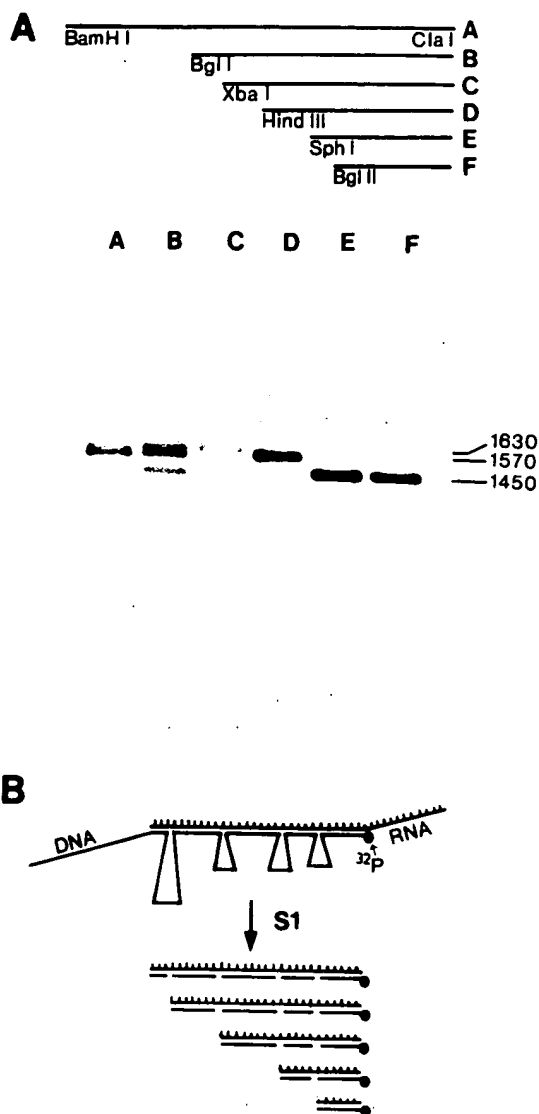


Fig. 2. (A) The DNA fragments used for hybridization to poly(A)<sup>+</sup> RNA are indicated above the autoradiogram. The DNA fragments carried the label at the 5' terminus of the left end. S1-resistant RNA/DNA hybrids were electrophoresed on 1.2% agarose gels. (B) A schematic drawing: formation of RNA/DNA hybrids, digestion with endonuclease S1 and possible S1-resistant hybrid molecules.

cated in Figure 3c. This position is within the interval between the *Xba*I and *Hind*III site which was determined above (Figure 2). The number of nucleotides between the last nucleotide of the cDNA that had been determined with certainty and the 3' end of the reverse transcript could be determined accurately, because random termination of the reverse transcriptase reaction caused a ladder of nucleotides in each lane. The distance was found to be 34 and 35 nucleotides. No dinucleotide AG, which could serve as an intron/exon boundary, is found in this interval within the DNA sequence. Thus the experiment places the transcription start tentatively, as shown in Figure 3.

A *Hind*III site is located within the first exon, as determined by the primer extension experiment. DNA was labeled at this *Hind*III site and used for a S1 protection experiment. Poly(A)<sup>+</sup> RNA obtained from wild-type endosperm protects DNA fragments of a length between 32 and 46 bp (Figure 3b). The four most pronounced RNA species are located at the positions indicated in Figure 3c. This confirms the transcription start site

determined by the reverse transcription experiment in general, but places the major transcription start sites 3–6 bp in front of the terminus of the reverse transcript. The reason for this is unknown. The most prominent signal detected in the S1 mapping experiment is defined as transcription start and numbered +1 in our sequence data. The boundary between exon 2 and intron 2 was placed from the known size of exon 2 (115 bp) and the GT/AG rule (Breathnach and Chambon, 1981). Approximately 40 bp upstream and downstream of this splice site (shown in Figure 3c) no other GT dinucleotide is found in the genomic sequence.

#### The insertion sites of *Ds* in the *sh-m5933* and *sh-m6233* alleles

The 4.1-kb insertion in the *sh-m6233* allele is flanked by the duplication of the 8-bp CTTGTCCC. This 8-bp sequence is found at positions 193–220 in intron 1. In the *sh-m5933* allele (Courage-Tebbe *et al.*, 1983), only the 3'-adjacent sequence of the 30-kb insertion is known. The 8-bp sequence GTCGCTTT located adjacent to the insertion site on the 3' side is located at positions 2970–2977 in intron 7. This insertion site is 12 bp upstream of the boundary to exon 8.

#### Protein coding capacity of the Shrunken mRNA

The first AUG is found in exon 2 and starts an open reading frame of 802 amino acid codons which is terminated by a TAG stop codon in exon 16. The predicted mol. wt. of the sucrose synthase monomer is 91.731 daltons. The predicted amino acid composition has been confirmed by a total hydrolysis of purified sucrose synthase protein.

#### Discussion

We have analyzed the transcription unit of the sucrose synthase gene and ~1.1 kb upstream of the transcription start by DNA sequencing. The results are discussed from 5' to 3' following the direction of transcription.

#### The 5'-flanking region of the sucrose synthase gene

We have examined the DNA sequence which extends ~1 kb upstream of the starting point of transcription, for sequences that might play a role in expression of the gene. Around position –610 we find a 16-bp direct repeat

TGGCGGGGA.GGAAATA  
TGG.GGGGAGGGAAATA

which is separated by 22 nucleotides. A tandem duplication of 15 bp TTTAGGAAAAA<sub>T</sub>TAG is found in front of a zein C gene, which is also expressed in the endosperm tissue (Langridge and Feix, 1983). This duplication is located between the two promoters used for transcription of the gene and oriented opposite to the direction of transcription. Comparison of the duplicated sequences reveals a conserved part GGAAA<sub>A</sub>A found in front of both endosperm genes. The sequence GGAAA found in all four repeats is also part of a consensus sequence found in animal enhancer elements TGGAAA (Weiher *et al.*, 1983).

The 22 nucleotides separating the duplicated 16 mers in front of the sucrose synthase gene contain a 4-fold repetition of the pentanucleotide GGTGG, a fifth copy overlaps the 16 mer duplicated downstream. It is interesting to note that the trinucleotide GTG is present five times within this region. This trinucleotide is reported to be frequent in procaryotic and eucaryotic DNA sites where interactions with proteins are anticipated for regulation or recombination (Cheng *et al.*, 1984).

Short nucleotide sequences have been implicated in the coordinate expression of eucaryotic genes (Davidson *et al.*, 1983). Therefore a possible role of these sequences in the expression

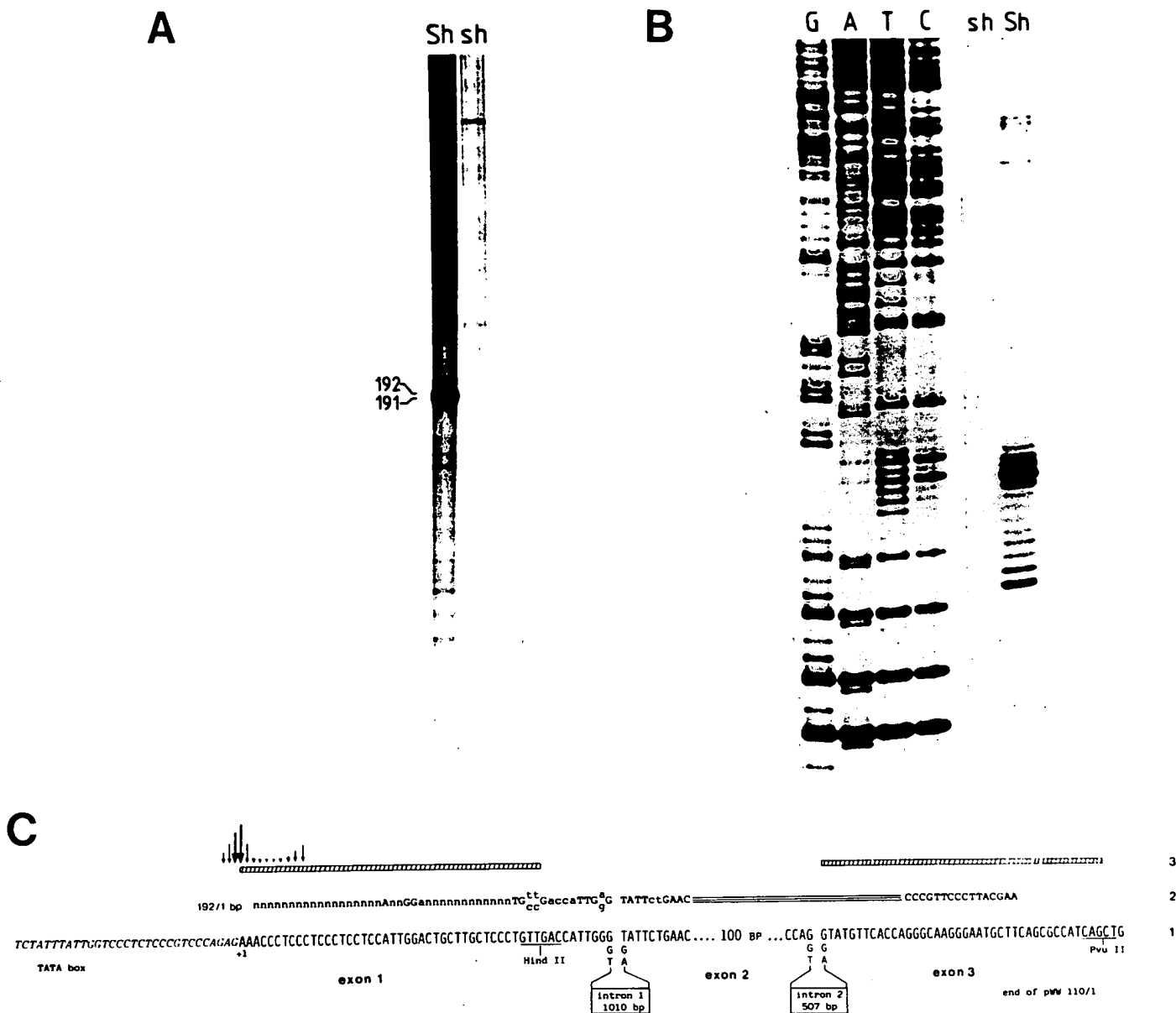


Fig. 3. (A) Primer extension experiment using the synthetic oligonucleotide from exon 3 as a primer for AMV reverse transcriptase [see (C)]. (B) Estimation of the transcription start by a S1 mapping experiment. The *Hind*III site shown in C was labeled at the 5' terminus and used for DNA sequencing and the RNA protection experiment. The RNA sources used in the experiments were isolated from kernels with *Sh* phenotype or from kernels homozygous for the deletion allele *sh bz-m4*. (C) Summary of the experiments which determine the transcription start. The DNA sequence (1) shows exon 1 and 2 and parts of exon 3, as well as some sequences upstream of the transcription start. Above in line 2 the sequence of the synthetic oligonucleotide used for priming of the reverse transcriptase, its position within exon 3 and the end of the reverse transcript shown in part A is demonstrated. The cDNA sequence information obtained at the exon 1/2 boundary is shown. The size of the letters is a measure of how unambiguously the base could be determined. Line 3 shows the size of DNA fragments obtained after S1 digestion of RNA/DNA hybrids. The main signal in the S1 mapping experiment is designated as major transcription start (+1).

of the sucrose synthase gene is presently investigated in a heterologous plant system (*Nicotiana tabacum*) as well as a transient expression experiment in maize protoplasts.

At position -124 upstream of the main transcription start the sequence CCAAGCT has similarity to the CAAT-box (GG<sub>2</sub>CAATCT, Breathnach and Chambon, 1981). The sequence (TATTTATT) located at position -29 resembles a TATA box. Both positions are marked in Figure 4.

#### The structure of the transcription unit

Most of the gene structure was deduced from comparison of a nearly full length cDNA clone to the genomic DNA sequence. The DNA sequence of the genomic clone has been determined

by sequencing both strands, while only one strand of the cDNA clone was analyzed. As the cDNA clone pWW110/1 ends in exon 3 (Figure 1), the position of exon 1 and 2 within the genomic sequence and the 5' end of exon 3 were determined by a combination of primer extension and S1 mapping experiments. At the transcription start site the results from these two types of experiments do not agree completely. The primer extension experiment places the transcription start at position +3 and +4. In the S1 experiment, start sites are found within a 13-bp region, extending from -3 to +10. Since these bands are found even after extended incubation with S1 (data not shown), they seem to really reflect different start sites rather than exonucleolytic digestion by S1. The two strongest start sites are located two nucleotides

**A PROMOTOR REGION**

-1081 CTCTTCATTCTTTTGTGTTTCCATGTCAGCGCCGATAGGCAGCTCTGTCTCTGTTTGGAGCCAAGCCAGCCCCAGCCCACACCTCGGGCAGCCCC  
 -981 GATGCGAGTGCCGTTTGGCGCCTCATCGCTACCGTTTGGGGGCTGCTCTGCTTCTGTTCTTCAAACGATGTCATGTCGCGCTGGGCAACTTT  
 -881 CTTGTGCGCCCTGTCGCTTGGCTGTGCTGACTGGACGCAGCTCCGGAGGTTTGGTTGTGCTTGGTTTTCTAGAGAACTGCCACTTGGCCGCCGCAC  
 -781 GTTCTTGGTGTTCGATCCTCCTCCTCACCGCTGTGCTCTGCGCAGGGCTTTTTCTGAGAGACCCATGTTTCTTTTACTTTTATAAACAGTTTACA  
 -681 TGCTATGTTTCTAGAAGGAGGGGAAACCTAATCCCCCTAATCCAATGGCGGGGAGGAAATAGGGTGGGGTGGGGTGGGGTGGGGGAGGGGAAATA  
 -581 TCTCGCTACTTTTTAATCCGACAAGTCATTTCGTTTGGCTCTGAATGATGATGACTGCAATGCTGATGCGACCTCGGGTGTGGATCACCGGCTTTT  
 -481 GGCTGCTCTCACCAATCAGCTGCAAGAAGATTAGAGCTCAAAAGAATTACAGAAAGAGAGCCTTTTTCTTTTCTTCTTGTGGGGTTCCTTTCAITTCG  
 -381 TGCTCTCTTTTCTCTGCCAGCCAGTCCGTCGCTCTTGGCTCCACTGCACCTGCACACAGGTACCCCGACCCGACTGTTCTAGACTCCATTAGAAAAA  
 -281 AAAAAGGTCTGAACCTTTCCGAAACCAGCCAGCCATTGGTCTGGCAGGCCAGCATATGCTAATGGAATTTTTTGCCGCATCATGAGTGGCCATCAGG  
 -181 ATTTGGAATTTCTGGTTTGTAGTAATACTGTAATTGGCAATTATCCATTGCCGAGTTTCAAGCTCCGTCAGCTGAACGTGGACCCCTACCATCTGCAC  
 -81 CAGCTCGGCACCTCACGCTCGCAGCGCATGGAGCCTAGGAGCAGCTGCCCGCTTATTATTGGTCCCTCTCCCGTCCAGAGAAACCTCCC +10

CAAT-box  
TATA-box

**B 3'-TERMINUS**

5541 CCTCTCGTTTTTTTCTTGTGTTGAGCGTTTTTGGGCAGCGCTGGCCTGGTTCCTAGTATGGTGGGAATGGCTGCACCTTTTGCTTCGATAATAAAATG  
 5641 CCTGCTCGTTACCTGTCTTCCAGAGTGCAATGCGATGTTCTGTGTCCAGGTCGTGTGTTCTGACTGATGGCGATGTGTGTTCTTCTGTTAATCGCC

polyadenylation  
signal  
AATAAA

polyA tail in pWW 110/1

Fig. 4. Selective DNA sequences. (A) DNA located in front of the transcription start. (B) 3' end of the sucrose synthase gene. The arrow 7 bp in front of the poly(A) tail in pWW110/1 indicates the position where polyadenylation starts in the cDNA clone isolated from W23 x K55.

upstream from those determined by the primer extension experiments. This difference may be explained by incomplete extension due to the cap structure at the 5' end of the mRNA (Banerjee, 1980; Nathans and Hogness, 1983). It is not clear why the microheterogeneity observed in the S1 protection experiment is not also seen in the primer extension experiment. It is possible that the minor bands have not been detected in the primer extension experiments due to the high background found in these autoradiographs. Microheterogeneity at the 5' end of the mRNA has also been determined for a soybean lectin gene (Vodkin *et al.*, 1983) as well as for the maize alcohol dehydrogenase gene (Dennis *et al.*, 1984).

The first ATG is found 72 nucleotides downstream of the main transcription start within exon 2. It is located within the sequence GAGCCATGG, which has extensive homology with other translation start sites of eucaryotic genes (CCACCATGG, Kozak, 1984).

This ATG starts an open reading frame of 2406 bp or 802 amino acids which predicts a mol. wt. for the protein of 91.731 daltons. The open reading frame is followed by an untranslated region of 269 nucleotides between the terminal TAG and the poly(A) addition in our cDNA clone pWW110/1. The exact polyadenylation site cannot be determined, because genomic clone and cDNA diverge at the position of two A residues, as is often found in polyadenylation sites. 31 nucleotides in front of the

poly(A) tail the sequence AATAAA is found. This signal is usually found in front of the poly(A) addition site (Proudfoot and Brownlee, 1976; Proudfoot, 1984). The sequence CAYUG is often observed between the AATAAA signal and the polyadenylation site (Berget, 1984). The only sequence slightly resembling this signal is the pentanucleotide CAGAG located three nucleotides in front of the poly(A) tail.

A cDNA clone isolated from maize line W23 x K55 by Chaleff *et al.* (1981) has a poly(A) tail starting 7 bp upstream of the poly(A) addition site in pWW110/1 (Sheldon *et al.*, 1983; C. Hannah, personal communication). The reason for the utilization of different polyadenylation sites remains unclear because the corresponding genomic sequences of W23 x K55 and line C are not known and might contain differences, which could explain this result. The sequence AATAAA is also found three times within the primary transcript, in introns 2, 6 and 11 (Table I). This confirms the observation by Proudfoot (1984) that AATAAA is not sufficient for poly(A) addition to occur or else that some poly(A) addition products are unstable and escape detection.

**Exons and introns**

The sizes of the exons and introns are shown in Figure 1. The exon-intron borders are listed in Table I. Only intron 1 (1114 bp) and 2 (511 bp) are larger than 162 bp, seven of the other introns are even shorter than 100 bp.

Table I.

Selected sequences in the transcription unit	
1080	GAGCCATGG translation start
1585	CTTCAGATCTAATAAAAAGGATATGAGATGCCATC
2592	GTGAATGCTCAATAAAAACGTTCTGTACTTGCTATGG
3852	TTTAGTAGTAATAAAAAGTATGTGATGTTTCT
5401	TAG end of the open reading frame

## Exon/intron boundaries

49	GGG	<u>GTATGCTT</u>	intron 1	AGTCGAAATGCGAG	TAT
1177	CAG	<u>GTGGGCTT</u>	intron 2	TACCACTTCTACAG	GTA
1809	CAG	<u>GTAACT</u>	intron 3	TTGCTGCATATAG	GAA
2048	ACA	<u>GTAAAGTTC</u>	intron 4	TCCTTTTACCAG	ATC
2320	ACG	<u>GTGAGCTT</u>	intron 5	GTTTCTGTTACAG	ACG
2589	TAG	<u>GTGAATGC</u>	intron 6	ATGATCTGTGTTAG	GTT
2904	CAG	<u>GTACAAAA</u>	intron 7	CAGTCGCTTGCAG	GTT
3075	ATT	<u>GTATGTTT</u>	intron 8	CTTATTGTTGCAG	GTT
3344	GAG	<u>GTATACAG</u>	intron 9	ATTCTGTGCTGCAG	GAT
3544	CAG	<u>GTCTGTTT</u>	intron 10	GTACATACTTGCAG	TGT
3820	AAG	<u>GTAGAATT</u>	intron 11	TGTTGTTTCTGCAG	CAA
4125	CAA	<u>GTGAGTAT</u>	intron 12	TTACTTGCTTCCAG	GTT
4578	CAG	<u>GTATATGC</u>	intron 13	TTTTGTGGGTAG	CCT
4951	GAA	<u>GTATGCAT</u>	intron 14	TTGGATTGCTCAG	GTA
5252	CTG	<u>GTAAAGCCG</u>	intron 15	TTTCTGGAATCCAG	GCA

The numbers in front of the short DNA sequences shown correspond to the number of the first nucleotide within our complete genomic sequence information. The positions of the potential polyadenylation signals AATAAA within the transcribed region are located in introns 2, 6 and 11, respectively.

The borders between exons and introns are in full agreement with the GT-AG rule (Breathnach and Chambon, 1981). The consensus sequence of introns at the donor site (Mount, 1982) is well conserved with the exception of position +6, where we observe a preference (seven out of 15) of C instead of T. The generally less-conserved acceptor site agrees well with the consensus sequence of published introns of animal genes.

Introns 1–14 carry a stop codon in frame, thus preventing translation of the unspliced RNA. Intron 15 is an exception. It could be translated in-frame with exons 15 and 16, yielding a translation product ending at a stop codon 67 bp in front of the AATAAA in a region that is not translated from the mature mRNA. Whether this is of biological significance is not known.

*Base composition*

Three contiguous exons, no. 13, 14 and 15, are significantly more GC-rich (57, 58 and 60%) than the other exons (51%). No such deviation is seen in the introns, which are in general more AT-rich than the exons. Much of the GC excess arises in third codon position and is thus not ascribable to the amino acid sequence. It will be interesting to see whether the 3-dimensional structure of the protein indicates that exons 13–15 form a separate domain. In this case, it could be discussed whether the gene is composed of two subgenes that have evolved separately. The lack of GC excess in the introns might then indicate that these have been added to the gene after the evolution of the coding domain. The dinucleotide CpG is under-represented in eucaryotic DNA (Bird, 1980). The distribution of CpG in the exons is shown in Figure 1. Again, exons 13, 14 and 15 show an excess of CpG, which is even higher than expected on the basis of the CG-content (46 expected/64 observed).

*Codon preferences*

Codon usage in the sucrose synthase gene has been tabulated and is available from the authors on request. An interesting deviation

Table II.

Transitions: 11	Transversions: 6
C→T (5), T→C (4)	A→C, G→T, C→G
A→G, G→A	C→A, T→G, T→A

Deviations between the DNA sequence of the genomic clone and the cDNA clone pWW110/1 within the protein coding region. The G to A transition is a difference between the genomic clone and the cDNA clone pKS500.

from a random usage is found for those codons where the second base is a T, and where A and G are used synonymously in the third position. In these cases, G is preferred strongly over A (valine GTG/A: 17/2; leucine TTG/A: 16/0; CTG/A: 34/3). A preference for the TG over TA in second and third codon position is also found in other eucaryotic genes (Beyreuther *et al.*, 1983) but to a lesser extent. The codon preference described results in a measurable though numerically smaller preference for TG over TA dinucleotides in the exons. No such preference is seen in the introns. This might indicate that the selection leading to this preference is exerted at the level of translation. It has not been reported, however, that these codons are served by different tRNA molecules (Sprinzl and Gauss, 1984).

*Evolutionary aspects*

The genomic clone sequenced by us was isolated from one of McClintock's strains. Our long cDNA clone pWW110/1 was isolated from line C which is a corn belt dent. Our short cDNA clone pKS500 is derived from the German commercial hybrid EDO. The genomic clone described by Sheldon *et al.* (1983) was isolated from Black Mexican Sweet. Some sequence information from the 3' end of the gene of this clone is available. We could thus search for sequence heterogeneities between these maize lines.

A comparison between the genomic and the cDNA sequence yields information about the exons. Since our cDNA clone (pWW110/1) is not complete at the 5' end and since we have not sequenced that part of it which was already known from the cDNA clone pKS500, we can compare a length of 2100 bp of exon DNA only. In the exon sequences investigated, we find 16 base substitutions, both transitions and transversions (see Table II). Though the cDNA clone has been sequenced from one strand only, we have inspected our gels carefully and could read the positions in question unambiguously. Whether the difference arises from mistakes of the reverse transcriptase cannot be excluded without sequencing another cDNA clone. It is noteworthy, however, that all of these differences are located in third codon positions and none of them leads to an amino acid substitution. This can hardly be explained by random mistakes of the reverse transcriptase reaction. Within the 270-bp protein coding region of our previously isolated cDNA clone pKS500 (Geiser *et al.*, 1980) we find one transition in the 3rd codon position, which is also silent on the protein level. Miyata *et al.* (1982) reported an accumulation of silent base substitutions in several animal genes of  $5.37 \times 10^{-9}$ /year.

We have compared a sequence of 2100 bp length between two maize lines, of which 540 positions potentially lead to silent substitutions. As we found 16 differences between the two sequences, we arrived at an evolutionary distance of  $6 \times 10^6$  years between the two maize lines implying a separation time of  $3 \times 10^6$  years. If we compare the 3'-untranslated region of the two sucrose synthase isolates mentioned above, we find within 270 bp 10 base substitutions (in addition to two deletion/insertion mutations). Since the 3'-untranslated region does not allow the distinction

GGTTCCTA <sup>G</sup> CTTCCTA GTA ... TGG	cDNA clone pWW 110/1
GGTTCCTA ..... GTA ... TGG	hypothetical precursor
GGTTCCTA ..... GTA TGG TGG	genomic clone

Fig. 5. Comparison of DNA sequences found in the 3'-untranslated region of different alleles of the *shrunk* locus, which could reflect transposon footprints. A possible common precursor allele is shown.

between silent and other codon positions, we compare all substitutions in this sequence of 270 bp with the number of substitutions in the 540 silent positions within coding sequences. The numbers found (16/540 and 10/270) are so similar that we must assume that the nucleotide exchanges in 3'-untranslated regions are accumulated at similar rates and are thus no more or less conserved than the silent positions within the coding region.

The calculated distance between the two alleles would mean that the two alleles have been separated  $3 \times 10^6$  years ago. This is two orders of magnitude longer than the time that has elapsed since the domestication of maize (Galinat, 1977). This may either mean that *Z. mays* has not gone through a very small population size upon domestication and that thus much of the diversity of its progenitors is reflected in its present genotype, or else that *Z. mays* evolves two orders of magnitude more rapidly than other species.

Comparison between these two sequences and those available from some other lines (Black Mexican Sweet, Sheldo *et al.*, 1983; pKS500, this study; W23 x K55, Chaleff *et al.*, 1981) detects a varying degree of divergence ranging from no substitutions between pKS500 and isolated from the German line EDO and the genomic clone sequenced here to 17 deviations between Black Mexican Sweet and line C. These results may indicate a lack of selection for the sequence of the 3'-untranslated region of the sucrose synthase gene.

In any case, it is interesting to note that we find no amino acid polymorphisms. It is surprising that a protein as large as sucrose synthase (mol. wt. 92 000) does not have amino acids, the replacement of which is neutral enough to be detected as a sequence polymorphism. This result is similar, however, to findings with the alcohol dehydrogenase gene of *Drosophila melanogaster* (Kreitmann, 1983) and supports his suggestion that most amino acid replacements are not neutral, but are rather selected against.

Spontaneous mutations in maize are often caused by the insertion of transposable elements. The integration of a transposon creates a short sequence duplication in the host sequence. In contrast to transposons of other organisms, these duplications persist upon excision of the transposon. They are, however, often mutated during the excision process (Sachs *et al.*, 1983; Bonas *et al.*, 1984; Fedoroff *et al.*, 1983; Weck *et al.*, 1984). Mutations occurring during the excision process are explained by a hypothesis advanced by Nevers *et al.* (1985).

If the integration of the transposon has occurred in an exon, the remaining duplication alters the predicted sequence of the revertant protein, if the number of remaining nucleotides is three or a multiple thereof. Two such instances have been described by Schwarz-Sommer *et al.* (1985). Proteins found in maize strains after the reversion of a transposon-induced mutation are sometimes altered in their properties (Echt and Schwartz, 1981; Dooner and Nelson, 1979; Shure *et al.*, 1983). These alterations may be due to the presumed insertions of one or a few additional amino acids.

Schwarz-Sommer *et al.* (1985) have also compared intron sequences of the *Waxy* gene of *Z. mays* obtained from different lines. They found perfect or near-perfect sequence duplications that could be ascribed to the insertion and subsequent excision of transposable elements. They advanced the hypothesis that transposon insertions and excisions are frequent events and may thus considerably contribute to the evolution of the maize genome. We have analyzed the 3'-untranslated sequences of five different *Sh* alleles and found two duplications of 7 and 3 bp each that are present in one and missing in the other strain (Figure 5). The two alleles may have evolved from a common precursor. The 7-bp duplication could be the footprint of an as yet unknown transposable element, or it could have been generated from an insertion causing a larger duplication of the insertion site by the deletion of one or a few base pairs (Bennetzen *et al.*, 1984; Döring and Starlinger, 1984). The 3-bp duplication could be the result of an En/Spm insertion and subsequent excision (Schwarz-Sommer *et al.*, 1985).

It must be asked, however, how much these transposon footprints contribute to the evolution of proteins. It is noteworthy in this respect that many of the transposon insertions known have occurred within an intron. This is also true for the three analyzed transposon insertions in the *Shrunk* gene. It is conceivable that many transposon insertions in exons leave footprints that are not comparable with protein functions. Such transposon insertions will not appear to revert frequently and may thus not be recognized as caused by transposable elements. The above-mentioned observations about amino acid replacements caused by single base substitutions supports this assumption. It should be remembered, however, that enzymes found presently may well be adapted to their function and that a more rapid evolution may have occurred in the past.

## Materials and methods

### Materials

The genomic clone SS1 (Geiser *et al.*, 1982) and the cDNA clone pKS500 (Geiser *et al.*, 1980) were described previously. pWW110/1 was isolated from a cDNA library prepared by Schwarz-Sommer *et al.* (1985) and isolated by standard methods (Maniatis *et al.*, 1982).

Restriction endonucleases were purchased from BRL (Neu-Isenburg), Biolabs (Dreieich) or Boehringer (Mannheim). T4 polynucleotide kinase, DNA polymerase large fragment (Klenow enzyme), alkaline phosphatase and DNase I from Boehringer (Mannheim); T4 ligase from Biolabs (Dreieich). Endonuclease S1 was from Sigma (München). AMV reverse transcriptase was from a preparation of J. Beard (Life Science).  $^{32}$ P-labeled nucleoside triphosphates were purchased from Amersham Buchler (Braunschweig).

### Subcloning and plasmid preparation

Vectors were pBR322 (Bolivar *et al.*, 1977), pUC9 (Vieira and Messing, 1982) and pUR250 (Rüther, 1982), which were used for transformation of *Escherichia coli* K12 strains HB101 or RR1ΔM15 (Rüther, 1982). Plasmid DNA was prepared in small amounts by the alkaline lysis method (Maniatis *et al.*, 1982) and larger quantities by the method of Clewell and Helinski (1969).

### RNA preparation

*Sh* RNA was isolated from the German hybrid EDO. The maize line carrying the allele *sh bz-m4* was obtained from B. McClintock. Immature ears were frozen in liquid nitrogen 20–22 days after pollination and stored at  $-70^{\circ}\text{C}$ . Frozen kernels were ground, incubated with proteinase K, phenol/chloroform extracted as described by Klopstech and Schweiger (1976). Poly(A)<sup>+</sup> RNA was isolated by oligo(dT)-cellulose affinity chromatography (Collaborative Research). RNA was stored in 70% ethanol at  $-70^{\circ}\text{C}$ .

### DNA sequencing

The chemical degradation method (Maxam and Gilbert, 1980) was used with modifications described by Garoff and Ansorge (1981). Ordered deletions were created by random DNase I cleavage within the maize insert and a unique restriction site in the vector (Frischauf *et al.*, 1990). Deletions with end points at distances of 250–300 bp were used for sequencing both strands after labeling by poly-

nucleotide kinase or fill in reaction by Klenow enzyme. Some restriction sites used for labeling were determined by sequencing DNA fragments spanning these sites.

#### Preparation of DNA-RNA hybrids and digestion with endonuclease S1

DNA-RNA hybrids were prepared as described by Berk and Sharp (1979). 5–10 µg of poly(A)<sup>+</sup> RNA or equivalent amount of tRNA were hybridized to 50 ng up to 1 µg of DNA fragments within a total volume of 20 µl. Hybridization was carried out overnight at 43°C. Single-stranded nucleic acids were digested with 200 U S1 endonuclease at 25°C for 30–60 min. Reaction was stopped by addition of 5 µg tRNA and 2.5 volumes ethanol.

#### Primer extension assay and RNA sequencing

The synthetic oligonucleotide AAGCATTCCCTTGCCC was synthesized by a 380 A DNA synthesizer (Applied Biosystems) with subsequent purification of the 16 mer on a 20% acrylamide gel. The reverse transcriptase reaction was carried out as described by Hamlyn *et al.* (1978) with slight modifications. 2.5 pmol oligonucleotide were incubated in 50 mM Tris pH 8.3, 150 mM KCl, 10 mM MgCl<sub>2</sub>, 20 mM DTT, 40 µM dTTP, dGTP, dCTP each and 6 µCi [ $\alpha$ -<sup>32</sup>P]ATP (400 Ci/mmol) with 2 U AMV reverse transcriptase at 42°C. After 15 min 1 µl of chase mix (250 µM dNTP in 100 mM Tris pH 8.3, 1 U AMV reverse transcriptase/µl) were added and incubation continued for 15 min. 2 µg tRNA were added, nucleic acids were ethanol precipitated and reverse transcription products analyzed on 6% polyacrylamide gels containing 7 M urea. For sequencing four reactions were carried out in the presence of one dideoxynucleotide each.

The computer analyses of DNA sequences were run on a VAX/VMS computer version 3.5. The computer programs used were developed at the University of Wisconsin (Devereux *et al.*, 1984).

#### Acknowledgements

We are greatly indebted to A.Gierl and Zs.Schwarz-Sommer, Köln-Vogelsang, for the opportunity to use an endosperm cDNA library from line C maize for the isolation of pWW110/1 clone. We thank H.Reinke and K.Beyreuther (Köln) for analysis of the purified sucrose synthase protein. This work was supported by Deutsche Forschungsgemeinschaft through SFB 74.

#### References

- Banerjee, A.K. (1980) *Microbiol. Rev.*, **44**, 175-205.
- Bennetzen, J., Swanson, J., Taylor, W.C. and Freeling, M. (1984) *Proc. Natl. Acad. Sci. USA*, **83**, 4125-4129.
- Berget, S.M. (1984) *Nature*, **309**, 179-181.
- Berk, A.J. and Sharp, P.A. (1977) *Cell*, **12**, 721-732.
- Beyreuther, K., Stüber, K., Bieseler, B., Bovens, J., Dildrop, R., Geske, T., Triesch, I., Trinks, K., Zaiss, S. and Ehrling, R. (1983) in Jensen, U. and Fairbrothers, D.E. (eds.), *Proteins and Nucleic Acids in Plant Systematics*, Springer-Verlag, Berlin/Heidelberg.
- Bird, A.P. (1980) *Nucleic Acids Res.*, **8**, 1499-1505.
- Bolivar, F., Rodriguez, R.L., Green, P.J., Betlach, M.C., Heynecker, H.L., Boyer, H.W., Crossa, H.J. and Falkow, S. (1977) *Gene*, **2**, 95-113.
- Bonas, U., Sommer, H. and Saedler, H. (1984) *EMBO J.*, **3**, 1015-1019.
- Breathnach, R. and Chambon, P. (1981) *Annu. Rev. Biochem.*, **50**, 349-383.
- Burr, B. and Burr, A.F. (1981) *Genetics*, **98**, 143-156.
- Chaleff, D., Mauvais, J., McCormick, S., Shure, M., Wessler, S. and Fedoroff, N. (1981) *Carnegie Inst. Wash. Yearbook*, **80**, 158-174.
- Cheng, S., Arnd, K. and Lu, P. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 3665-3669.
- Chourey, P.S. (1981) *Mol. Gen. Genet.*, **184**, 372-376.
- Chourey, P.S. and Nelson, O.E. (1976) *Biochem. Genet.*, **14**, 1041-1055.
- Clewell, D.B. and Helinski, D.R. (1969) *Proc. Natl. Acad. Sci. USA*, **62**, 1159-1166.
- Courage-Tebbe, U., Döring, H.P., Fedoroff, N. and Starlinger, P. (1983) *Cell*, **34**, 383-393.
- Davidson, E.H., Jacobs, H.T. and Britten, R.J. (1983) *Nature*, **301**, 468-470.
- Dennis, E.S., Gerlach, W.L., Pryor, A.J., Bennetzen, J.L., Inglis, A., Llewellyn, D., Sachs, M.M., Ferl, R.J. and Peacock, W.J. (1984) *Nucleic Acids Res.*, **12**, 3983-4000.
- Devereux, J., Haeberli, P. and Smithies, O. (1984) *Nucleic Acids Res.*, **12**, 387-395.
- Döring, H.P. and Starlinger, P. (1984) *Cell*, **39**, 253-259.
- Döring, H.P., Geiser, M. and Starlinger, P. (1981) *Mol. Gen. Genet.*, **184**, 377-380.
- Dooner, H.K. and Nelson, O.E. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 2369-2371.
- Downton, W.J.S. and Hawker, J.S. (1973) *Phytochemistry*, **12**, 1551-1556.
- Echt, C.S. and Schwartz, D. (1981) *Genetics*, **99**, 275-284.
- Fedoroff, N., Mauvais, J. and Chaleff, D. (1983) *J. Mol. Appl. Genet.*, **2**, 11-30.
- Frischauf, A.M., Garoff, H. and Lehrach, H. (1980) *Nucleic Acids Res.*, **8**, 5541-5549.
- Galinat, W.C. (1977) in Sprague, G.F. (ed.), *Corn and Corn Improvement*, American Society of Agronomy Inc., Madison, WI, pp. 1-47.
- Garoff, H. and Ansorge, W. (1981) *Anal. Biochem.*, **115**, 450-457.
- Geiser, M., Döring, H.P., Wöstemeyer, J., Behrens, U., Tillmann, E. and Starlinger, P. (1980) *Nucleic Acids Res.*, **8**, 6175-6188.
- Geiser, M., Weck, E., Döring, H.P., Werr, W., Courage-Tebbe, U., Tillmann, E. and Starlinger, P. (1982) *EMBO J.*, **1**, 1455-1460.
- Hamlyn, P.H., Brownlee, G.G., Chen-Chi Cheng, Gait, M.J. and Milstein, C. (1978) *Cell*, **15**, 1067-1075.
- Hawker, J.S. (1971) *Phytochemistry*, **10**, 2313-2322.
- Klopptech, K. and Schweiger, G. (1976) *Cytobiologie*, **13**, 394-400.
- Kozak, M. (1984) *Nucleic Acids Res.*, **12**, 857-872.
- Kreitman, M. (1983) *Nature*, **304**, 412-417.
- Langridge, P. and Feix, G. (1983) *Cell*, **34**, 1015-1022.
- Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, published by Cold Spring Harbor Laboratory Press, NY.
- Maxam, A. and Gilbert, W. (1980) *Methods Enzymol.*, **65**, 499-560.
- McCormick, J., Mauvais, J. and Fedoroff, N. (1982) *Mol. Gen. Genet.*, **187**, 494-500.
- Miyata, T., Hayashida, H., Kikuno, R., Hasegawa, M., Kobayashi, M. and Koike, K. (1982) *J. Mol. Evol.*, **19**, 28-35.
- Mount, S.M. (1982) *Nucleic Acids Res.*, **10**, 459-472.
- Nathans, J. and Hogness, D.S. (1983) *Cell*, **34**, 807-814.
- Nevers, P., Shepherd, N. and Saedler, H. (1985) *Adv. Bot. Res.*, in press.
- Preiss, J. and Levi, C. (1980) in Preiss, J. (ed.), *The Biochemistry of Plants*, Vol. 3, Academic Press, NY, pp. 371-424.
- Proudfoot, N. (1984) *Nature*, **307**, 412-413.
- Proudfoot, N. and Brownlee, G.G. (1976) *Nature*, **263**, 211-214.
- Rüther, U. (1982) *Nucleic Acids Res.*, **10**, 5765-5772.
- Sachs, M.M., Peacock, W.J., Dennis, E.S. and Gerlach, W.L. (1983) *Maydica*, **28**, 289-302.
- Schwarz-Sommer, Z., Gierl, A., Coyper, H., Peterson, P. and Saedler, H. (1985) *EMBO J.*, **4**, 579-583.
- Sheldon, E., Ferl, R., Fedoroff, N. and Hannah, L.C. (1983) *Mol. Gen. Genet.*, **190**, 421-426.
- Shure, M., Wessler, S. and Fedoroff, N. (1983) *Cell*, **35**, 225-233.
- Sprinzl, M. and Gauss, D.H. (1984) *Nucleic Acids Res.*, **12**, 1-58.
- Tsai, C.Y. (1974) *Phytochemistry*, **13**, 885-891.
- Vieira, J. and Messing, J. (1982) *Gene*, **19**, 259-268.
- Vieweg, G.H. (1974) *Planta*, **116**, 347-359.
- Vodkin, L.O., Rhodes, R.R. and Goldberg, R.B. (1983) *Cell*, **34**, 1023-1031.
- Weck, E., Courage-Tebbe, U., Döring, H.P., Fedoroff, N. and Starlinger, P. (1984) *EMBO J.*, **3**, 1713-1716.
- Weier, H., König, M. and Gruss, P. (1983) *Science (Wash.)*, **219**, 626-631.

Received on 13 March 1985

# Maize sucrose synthase-1 promoter directs phloem cell-specific expression of *Gus* gene in transgenic tobacco plants

(tissue specificity/glucuronidase/plant promoter/anaerobic induction)

N.-S. YANG\* AND D. RUSSELL

Department of Genetic Engineering, Agracetus, 8520 University Green, Middleton, WI 53562

Communicated by C. C. Tan, March 7, 1990

**ABSTRACT** Tobacco plants were transformed by using a chimeric gene construction, in which a corn sucrose synthase-1 gene (*Sh*) promoter was used to direct expression of the  $\beta$ -glucuronidase (*Gus*) reporter gene. Expression of *Sh-Gus* activity in these plants was found to be cell type specific. GUS activity was detected only in the phloem cells but not in any other cell types of vegetative tissues. In addition, *Sh-Gus* expression was found to be anaerobically inducible in tobacco roots. *Sh-Gus* was also expressed at high levels in the endosperm tissue of maturing tobacco seeds. We thus demonstrated that the corn *Sh* promoter can direct cell-type-specific and inducible expression in a heterologous dicotyledonous plant.

Corn sucrose synthase 1 (*Sh*) gene, encoding sucrose synthase 1 isozyme (UDPGlucose:D-fructose 2- $\alpha$ -D-glucosyltransferase, EC 2.4.1.13), is one of the few monocotyledonous plant genes whose biochemical function has been identified (1) and whose promoter sequence has been cloned (2–4). As sucrose synthase 1 isozyme is the primary enzyme of the various sucrose synthase isozymes, it is believed that it may play some important physiological role(s). The *Sh* gene was isolated from the corn *shrunk-1* locus, which is located on chromosome 9S (5). Enzyme activity studies with tissue extracts have shown that the sucrose synthase 1 isozyme in corn is expressed at high levels in endosperm, at a much reduced level in root, but not in green tissues or pollen (see ref. 6 for review). These results indicate that the *Sh* gene in corn is expressed in a tissue- or organ-specific fashion. Physiological studies have shown that *Sh* gene expression is anaerobically inducible in root and shoot tissues of etiolated corn seedlings (6, 7).

Using a chimeric gene construction, in which a corn *Sh* gene promoter is used to control the expression of the  $\beta$ -glucuronidase (*Gus*) gene, we transformed tobacco plants and examined in detail the expression of *Sh-Gus* activity in transgenic tissues. The observed cell-specific and inducible expression suggests the possible use of the *Sh* promoter for targeting chimeric gene expression to specific tissues in transgenic plants.

## MATERIALS AND METHODS

**Construction of Promoter Genes and Vectors.** The maize sucrose synthase promoter and first intervening sequence was isolated from Pvu55, a maize *Pvu* I genomic fragment in pBR327 provided by B. Burr (2). The *Pvu* I site at –1347, relative to the start of transcription, was converted to an *Xho* I site by linker mutagenesis (W. Swain, personal communication). A 2430-base-pair (bp) *Xho* I/*Nco* I fragment, containing the *Sh* promoter, first intron, and translation start site

was isolated and inserted into a promoter-less *Gus* expression vector (8). The resulting plasmid, designated pShGus, contains the *Sh* promoter from –1347 through the transcription start site, the first intervening sequence in the transcribed but nontranslated region, and the *Sh* gene translation start site at the *Nco* I site. pShGus was then cut with *Xho* I and inserted into an *Xho* I cut pTV4 vector (pTV4 contains a *nptII* gene for kanamycin selection and Ti border sequences; ref. 9) and was designated pTVShGus.

A soybean *rbcS* promoter was isolated from pSRS2.1 provided by R. Meagher (10). The *rbcS* promoter was altered to place an *Xho* I linker at about –1550 relative to the start of transcription and an *Nco* I site was placed at the translation start site at +45. The resulting 1595-bp *Xho* I/*Nco* I soybean *rbcS* promoter fragment was then inserted into a promoterless pCMC1100 vector at the *Xho* I/*Nco* I sites to form a soybean *rbcS-Gus-nos* 3' gene fusion designated pCMC2100. pCMC2100 was then cut with *Xho* I and the entire plasmid was inserted into pTV4 as described above. A soybean heat shock (*hs*) promoter was cloned with the gene *hs6871* described by Schoffl *et al.* (11). This promoter was cloned by using polymerase chain reaction (PCR) primers to amplify the promoter fragment from soybean genomic DNA and resulted in an *Xho* I site at about –420 and an *Nco* I site (+100) at the translation start site. The 520-bp *Xho* I/*Nco* I promoter fragment was transferred to a *Gus* expression vector and then to a transfer vector as described above for pTVShGus.

**Plant Transformation.** We generated tobacco plants with the promoter expression DNAs described above both via *Agrobacterium* and electroporation. Tobacco stem sections were transformed with disarmed Ti plasmid via *Agrobacterium* infection as described (9).

Alternatively, tobacco mesophyll protoplasts were electroporated by standard cell culture manipulations (12). Specifically, we used the following conditions: 10  $\mu$ g of plasmid DNA per ml, 200  $\mu$ g of salmon sperm DNA per ml (sonicated),  $2 \times 10^6$  cells per ml, 350 V/cm, 500  $\mu$ F, 25 msec using a PDS (Prototype Design Services, New York) electroporation apparatus. Transformed cells or tissues were selected on medium containing kanamycin (100  $\mu$ g/ml) and plants were regenerated (9). Transformed plants were grown under standard greenhouse conditions. Unless otherwise indicated, 2- to 3-foot-tall plants were used for assay of cell-type-specific expression of GUS activity.

**Histochemical Staining of GUS Activity.** Stem, leaf, root, and other parts of plant organs were cut from transformed plants. Thin (0.3–0.5 mm), cross tissue sections were hand cut and submerged in GUS reaction mixture as described (13). Tissue sections in GUS reaction buffer were vacuum infiltrated at 300 mmHg for 2 min. Tissues were incubated at 30°C for 1–8 hr, depending on the level of GUS activity expressed in test samples. Unless otherwise indicated, *Sh-Gus* tissues were stained for 8 hr, while *rbcS-Gus* and *hs-Gus*

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: PCR, polymerase chain reaction.  
\*To whom reprint requests should be addressed.



tissues were stained for 1–2 hr. The reaction was stopped by fixing in phosphate-buffered formalin, and chlorophyll-containing green tissues were bleached with 75% (vol/vol) ethanol. Cellular location of GUS activity was observed by light microscopy. Photographs were taken with a Zeiss photomicroscope and Kodak Gold 100 film.

**Anaerobic Induction and Heat Shock Treatment.** Anaerobic induction of root *Sh-Gus* activity was performed as follows.  $R_1$  progeny seeds of *Sh-Gus* transformed tobacco plants were grown for 4 weeks in the greenhouse (20 cm tall). Segments of main root (connecting to stem,  $\approx 0.5$  cm diameter) were cut off the plants and sliced longitudinally along the root axis to give two corresponding half-root samples. One set of half-root from each plant was submerged deep into 50 ml of MS medium in a 50-ml conical culture tube. Carbenicillin (100  $\mu$ g/ml) and nystatin (100 units/ml) were added to the MS medium (14) to prevent growth of microorganisms. The submerged half-roots were vacuum infiltrated (300 mmHg; 2 min) to provide better anaerobic conditions for the test. The second set of half-roots was rinsed twice in MS medium and placed with the cut side down in a humidified Petri dish. After 2 days, tissue sections were cut and GUS activities in the corresponding two half-root samples from each plant were compared. For heat shock treatment, freshly cut stem and leaf segments (1 cm long) were submerged in MS medium (14) supplemented with 0.1 M glucose. Stem and leaf tissues were then transferred to the same medium preheated to 45°C and incubated at 45°C in a water bath for 5 min with occasional shaking. Test tissues were then transferred to a 25°C medium, held for 10 min, and heat shock treated again as described above. Heat shocked and control segments were then held in a humidified Petri dish for 16 hr before sectioning and assaying for GUS activity.

**PCR Analysis.** DNA from  $\approx 10$  mg of plant tissue was extracted, ethanol precipitated, and redissolved. One hundred nanograms of these crude DNA preparations and 20 pmol of each of the appropriate primers were used under standard PCR conditions in a Perkin-Elmer thermal cycler as described (15).

## RESULTS

A simplified map showing the main features of corn *Sh* promoter, soybean *rbcs* promoter, and soybean *hs* promoter used in our *Gus* gene constructions is depicted in Fig. 1. The transformed tobacco plants and the  $R_1$  progeny were grown in the greenhouse, various organs were harvested, and thin tissue sections were stained for GUS activity and examined by light microscopy.

Fig. 2A shows the expression of GUS activity observed in a young root section of *Sh-Gus* transformed tobacco plants. Clusters of cells that were stained blue were localized in phloem tissues. No activity was detected in xylem, pith, cortex, or epidermal tissues. At high magnification (Fig. 2B), the cellular localization of GUS activity was found to be highly specific. A group of four to six cells residing at the center of phloem tissue clusters are highly stained, and the immediate parenchyma cells show little or no stain. The remaining cell types of the vascular cylinder, including xylem fiber, xylem parenchyma cells, pith parenchyma cells, pericycle, and endodermis cells were not stained. At still higher magnification (data not shown), we observed that two to three pairs of sieve elements and companion cells were localized at the center of each phloem tissue bundle, and both cell types were similarly stained with high GUS activity. The surrounding parenchyma cells of the phloem tissues, as shown in Fig. 2B and D, express little or no activity. At this level of cellular resolution, we thus have used the term "phloem cells" (including only sieve elements and companion cells) rather than phloem tissues to describe our results.

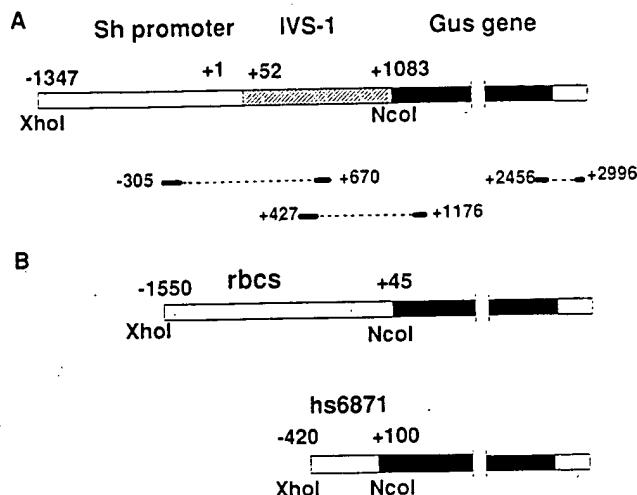


FIG. 1. (A) Maize *Sh-Gus* expression cassette. The location of the maize *Sh* promoter up to the translation start site at the *Nco* I site is indicated by the open box. The site of transcription initiation is indicated by +1. The first intervening sequence (IVS-1), which falls in the nontranslated leader sequence, is indicated by the hatched box. The *Gus* gene (solid box, truncated) and the *nos* 3' poly(A) termination region are also shown. For PCR analysis, synthetic oligonucleotides were used as primers. The locations of the primers are shown by the solid bars (-305; DR140, +670; DR139, +427; DR72, +1176; DR57, +2456; KB71, +2996). The region of the *Sh-Gus* gene amplified by PCR is shown by a dotted line (see also Fig. 3). (B) Constructs of the upstream region for soybean *rbcs* and soybean *hs* promoters. Promoter fragments were ligated to the *Gus* coding sequence as transcriptional fusions.

Fig. 2C shows the expression of GUS activity observed in a stem section of *Sh-Gus* transformed tobacco plant. A special feature of tobacco stem is that it contains two groups of phloem tissues—namely, the internal and external phloems, which are located along the two sides of the xylem. Most other dicotyledonous plants only have external phloem. As shown in Fig. 2C, GUS activity is detected in both types of phloem cells but not in any other cell types of the stem, including epidermis, xylem, and ground tissue cells of pith and cortex. At higher magnification (Fig. 2D), the cellular localization of GUS activity was found again to be highly specific. The phloem cells were stained, whereas the adjacent phloem parenchyma cells and vascular cambium cells were not. Phloem cell-specific expression of *Sh-Gus* was also observed in leaf, flower, and fruit tissues. Fig. 2E shows that only phloem cells in midrib and lateral veins of leaf tissue expressed high levels of GUS activity, whereas no activity was detectable in other leaf cell types. Phloem cell-specific expression of *Sh-Gus* activity was also observed in flower and fruit tissues (Fig. 2F and G, respectively).

In corn, the *Sh* gene is expressed at high levels in the endosperm of the immature kernel (1, 6). We have tested *Sh-Gus* expression in the endosperm of transgenic tobacco plants. Fig. 2G and H shows that GUS activity could be readily detected in the thin endosperm tissue of immature tobacco seeds. In dicotyledonous plants, endosperm tissue in seed is formed after fertilization as a thin layer of nutritive tissue aligned in a concave fashion with the inner seed coat (16). The tissue is initially transparent and is later either absorbed by the developing embryo or pressed in between the cotyledon and seed coat in the mature seed. A tissue section was cut longitudinally through a maturing seed (Fig. 2G). It shows that a high level of *Sh-Gus* activity is expressed in tobacco endosperm tissue, but there is no activity in the cotyledon (Fig. 2G). Fig. 2H shows that, in a younger immature seed and at high magnification, GUS activity in young endosperm is readily detected at the cellular level. To



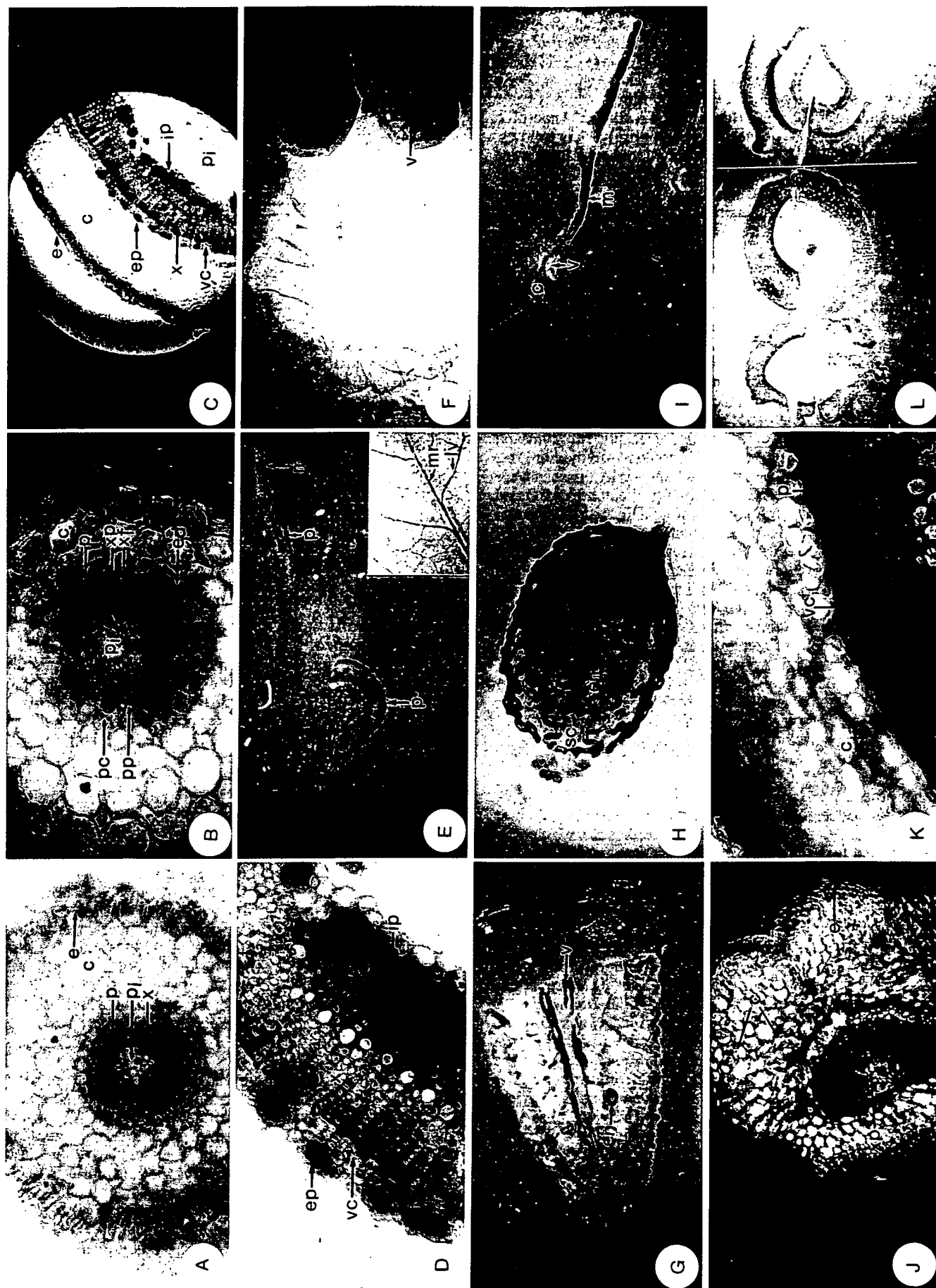


FIG. 2. (Legend appears at the bottom of the opposite page.)

better show the endosperm tissue, we cut the immature seed longitudinally into halves and gently teased out the embryo and cotyledon from endosperm and seed coat tissues. The focal plane of Fig. 2H was adjusted close to the bottom of the sliced half seed. In this case, the endosperm tissue was two to three cell layers thick and was lined with the seed coat. GUS activity was detected in small round endosperm cells but not in large polygonal seed coat cells. When *Sh-Gus* tobacco seeds were germinated in the dark for 4 days, enzyme activity could be detected only in the very fine veins extending through the root, hypocotyl, and cotyledon tissue, but not in any other tissues of the seedlings (data not shown). Hence, the pattern of phloem cell-specific expression of *Sh-Gus* activity, observed for mature transgenic tobacco plants, was conserved in young tobacco seedlings.

The same pattern of cell-specific expression for *Sh-Gus* in transgenic tobacco plants was observed for three independently transformed plants. Similar expression patterns were observed whether the plants were produced by electroporation or by *Agrobacterium* infection. In addition to regenerated  $R_0$  plants, the selfed  $R_1$  progeny (five plants for each transformant) were similarly tested for their GUS expression patterns and the same results were obtained as shown above. We therefore demonstrated that GUS expression dictated by the maize *Sh* promoter in transgenic tobacco plants was transmitted to the next generation.

Results showing localized GUS staining (Fig. 2 A–G) indicate that the maize *Sh* promoter directs phloem cell-specific expression of the *Gus* gene in transgenic tobacco plants. However, it is important to show that the results are not due to differential staining anomalies. For comparison, transgenic tobacco plants transformed with the same *Gus* gene but with different plant gene promoters (see Fig. 1) were tested under identical conditions for the GUS assay. Fig. 2I shows that tobacco plants transformed with soybean *rbcs* promoter express high levels of GUS activity in leaf mesophyll cells but not in phloem or other cell types of leaf tissues. Fig. 2J shows that, 16 hr after heat shock, tobacco plants transformed with the soybean heat shock promoter express high levels of GUS activity in virtually all cell types present in leaf tissues, including cortex, vascular cambium, xylem parenchyma, epidermis, and phloem cells. In stem, vascular cambium and phloem cells are apparently more responsive to the heat shock treatment than other cell types and were readily stained with GUS activity (Fig. 2K; see Fig. 2B for comparison). These results confirm that GUS enzyme produced in various cell types of tobacco tissues, including phloem, vascular cambium, xylem parenchyma, cortex, and mesophyll tissues, can be nondiscriminatively detected by our GUS activity assay. The experiments confirm that the phloem cell expression pattern observed for *Sh-Gus* transgenic plants is highly specific, is directed by the *Sh* promoter, and is not due to some unknown factors that may differentially affect activity staining in our GUS assay.

Expression of the *Sh* gene in young corn seedlings can be induced under anaerobic conditions (6). Springer *et al.* (7) have shown that the level of *Sh* gene transcripts increases 10 and 20 times in shoot and root, respectively. In *Sh-Gus* tobacco plants, we have observed that GUS activity is anaerobically inducible in root tissue. Using replicate half-root samples prepared from the root of the same plants, we

showed (Fig. 2L) that root GUS activity is clearly induced under anaerobic conditions. With the short time period ( $\approx 30$  min) used for GUS staining in this case, only a low level of activity was detected in the control (aerobic) root tissues, but high levels of enzyme activity were observed for the anaerobically treated half-root tissues. The expression of GUS enzyme remained phloem specific. When six  $R_1$  progeny from the same transgenic parent were tested for anaerobic induction of *Sh-Gus* expression in root, four responded with strong induction (similar to Fig. 2L). The other two plants showed little observable differences in GUS activity in half-root samples that were treated with or without anaerobic induction conditions, suggesting that these plants were only weakly induced above the basal expression level found in roots. Overall, our results suggest that *Sh-Gus* activity in root phloem cells of transgenic tobacco plants is anaerobically inducible and the *Sh* promoter is apparently responsible.

PCR analysis was performed to provide molecular evidence that an intact *Sh-Gus* gene was introduced into tobacco plants. Fig. 1 indicates the positions of the primers used to detect the presence of each of the regions of the *Sh-Gus* gene. Primers DR140 and DR139 detected a 975-bp fragment spanning the transcription start site from  $-305$  in the promoter to  $+670$  in the nontranslated intron. Primers DR138 and DR72 detected a 749-bp region from within the intron to the 5' coding region of the *Gus* gene. Primers DR57 and KB71 could detect a 540-bp region from the 3' end of the *Gus* coding region into the *nos* 3' terminator. These three sets of PCRs demonstrated that the promoter, intron, and *Gus* gene were intact. Fig. 3 shows the PCR results for two transgenic plants. Tobacco plant 3487 was produced via *Agrobacterium* infection, while tobacco plant 2713-14 is the  $R_1$  progeny of plant 2713, which was produced via electroporation. The data demonstrated that transgenic plants recovered from different transformation methods contain intact genes and show the same expression patterns.

## DISCUSSION

In this report, we demonstrated that the corn *Sh* gene promoter can direct highly specific, phloem cell-type expression of the *Gus* gene in transgenic tobacco plants. *Sh-Gus* was also found to express at high levels in endosperm tissue of immature tobacco seeds. In addition, the corn *Sh* promoter could be anaerobically induced in phloem cells of tobacco root tissue. Thus, the *Sh* promoter can target phloem cells as the tissue for foreign gene expression in genetically engineered plants. As phloem tissues of various plant organs are the target tissues for many plant pathogens, highly specific phloem expression of pathogen-resistant gene(s) via the *Sh* promoter may be utilized in disease control. It can also be used for anaerobic induction of foreign genes in root phloem cells.

Springer *et al.* (7) have demonstrated that expression of *Sh* gene at the transcriptional level is readily detectable in shoots of etiolated corn seedlings (including first and second leaves). A very low, but detectable, transcript level is found in mature green leaves. Upon anaerobic stress of young seedlings, the level of *Sh* transcripts increases 10–20 times in shoot and root, respectively. Our results in transgenic tobacco suggest that the expression of the *Sh* promoter is best described not

Fig. 2 (on opposite page). Histochemical localization of GUS activity in transgenic tobacco tissues. Expression of *Sh-Gus* gene in tobacco tissues: (A) Root at low magnification. ( $\times 4$ .) (B) Root at high magnification. ( $\times 15$ .) (C) Stem. ( $\times 4$ .) (D) Stem. ( $\times 15$ .) (E) Leaf midrib and mesophyll (Inset, part of a whole leaf). (F) Petal and ovary base. (G) Fruit and seed endosperm. (H) Immature seed endosperm. Expression of *rbcs-Gus* gene in tobacco tissue: (I) Leaf midrib and mesophyll. Expression of *hs-Gus* gene in tobacco tissues: (J) Leaf. (K) Stem. Anaerobic induction of *Sh-Gus* in root tissue of transgenic tobacco: (L) Half-root samples. (Upper) Anaerobic roots. (Lower) Control (aerobic roots). p, Phloem; ip, internal phloem; ep, external phloem; pp, phloem parenchyma; c, cortex; pi, pith; x, xylem; xp, xylem parenchyma; xf, xylem fiber; e, epidermis; v, vascular bundles; vc, vascular cambium; ed, endodermis; mr, midrib; lv, lateral vein; m, mesophyll; en, endosperm; sc, seed coat layers.

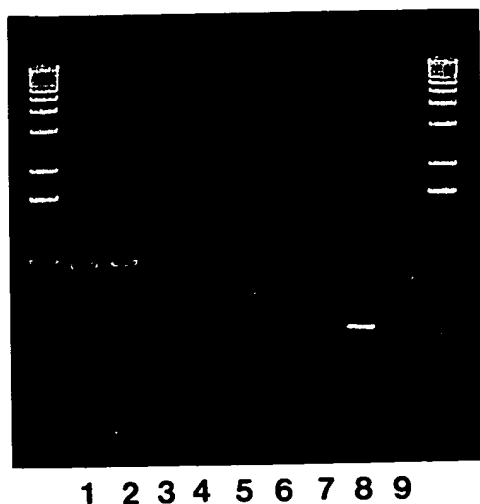


FIG. 3. PCR analysis to detect the *Sh-Gus* gene in transgenic plants. PCR was used to show that intact genes were transferred to the transgenic tobacco plants. Lanes 1–3, PCRs containing primers DR140 + DR139 and genomic DNA from Tob 3487, Tob 2713-14, and Havana 425 (nontransformed control), respectively. Lanes 4–6, reactions containing DR138 + DR72 and genomic DNA from Tob 3487, Tob 2713-14, and Havana 425, respectively. Lanes 7–9, reactions with DR57 + KB71 and genomic DNA from Tob 3487, Tob 2713-14, and Havana 425, respectively. Outer lanes are size markers.

at the organ level but at the cellular level. The *Sh* promoter is specifically expressed in phloem cells and it is the relative abundance of phloem cells in different organs that contributes to the organ-specific pattern one would observe by measuring *Sh* promoter activity in leaves, shoots, or roots. If one takes into account the relative abundance of phloem cells in these various organs, the pattern of expression we see in this system appears to be consistent with the expression observed in corn by Springer *et al.* (7). As shown in Fig. 2E, *Sh-Gus* expression is observed in leaf phloem cells. Since these cells account for <5% of the midrib cells and 1% of the leaf blade cells (estimated by microscopy), the leaf as an organ would be expected to show only low levels of *Sh* promoter expression. Therefore, when RNA, protein, or enzyme specific activities are measured from whole organ extracts, highly specific cell-type expression of a test gene could be drastically reduced or become nondetectable. Furthermore, efficiencies for tissue extraction are often found to vary considerably among different plant organs, this could further interfere with the biochemical assays for gene expression that is highly specific at the tissue or cell level.

Rowland *et al.* (17) showed very recently that the *Sh* gene transcripts are expressed in a cell-type-specific manner in corn root tissues. In their study, only the root tissues were tested. Expression was detected mainly in the vascular tissues, to a lesser extent in pith and epidermis, and not at all in cortex tissues. We have demonstrated that the *Sh-Gus* activity in transgenic tobacco is expressed only in the phloem cells of vascular tissues, but not in xylem parenchyma,

vascular cambium, or any other cell types of vegetative tissues. The differences observed for the *Sh* promoter expression in the roots of corn and tobacco may be due to differences inherent in the expression of the corn *Sh* promoter in the two species. Alternatively, they may be due to differences in the resolution of the different assay systems used. A direct comparison using the same assay systems or access to *Sh-Gus* transgenic corn would address these differences.

At present, the physiological role(s) of the sucrose synthase enzyme(s) in vegetative tissues of plants is not clear (7). Our observation on cell-type expression of the *Sh* promoter suggests that a physiological role of sucrose synthase activity may be specifically associated with the phloem tissues of plants—e.g., sucrose loading and transport through the phloem cells in vascular plants.

The fact that the *Sh-Gus* gene is effectively expressed in tobacco phloem cells suggests that both the transcriptional and translational apparatus in tobacco phloem are capable of recognizing the signals of corn *Sh* promoter and its intron sequences and can effectively utilize them to dictate the induction and cell-type-specific expression of the *Gus* coding sequences.

We thank Dr. Ben Burr for his generous gift of the *Sh* genomic clone and Dr. W. Swain for subcloning the *Sh* promoter fragment. We also thank Jim Fuller, Barry Cohen, and Joe Burkholder for technical assistance. We thank Drs. Oliver Nelson, Ray Evert, Judy Croxdale, and Paul Christou for many useful discussions.

1. Chourey, P. S. & Nelson, O. E. (1976) *Biochem. Genet.* **14**, 1041–1055.
2. Burr, B. & Burr, F. (1982) *Cell* **29**, 977–986.
3. Geiser, M., Weck, E., Doring, H. P., Werr, W., Courage-Tebbe, U., Tillmann, E. & Starlinger, P. (1982) *EMBO J.* **1**, 1455–1460.
4. Coe, E. H., Neuffer, D. A. & Hoisinton, D. A. (1988) in *Corn and Corn Improvement*, eds. Sprague, G. & Dudley, J. (Am. Soc. Agron., Madison, WI), pp. 81–258.
5. Werr, W., Frommer, W.-B., Maas, C. & Starlinger, P. (1985) *EMBO J.* **4**, 1373–1380.
6. Bailey-Serres, J., Kloeckener-Gruissem, B. & Freeling, M. (1987) *Plant Cell Environ.* **11**, 351–358.
7. Springer, B., Werr, W., Starlinger, P., Bennett, D. C., Zokolica, M. & Freeling, M. (1986) *Mol. Gen. Genet.* **205**, 461–468.
8. McCabe, D., Swain, W., Martinell, B. & Christou, P. (1988) *Biotechnology* **6**, 923–926.
9. Barton, K., Whiteley, H. & Yang, N.-S. (1987) *Plant Physiol.* **85**, 1103–1109.
10. Berry-Lowe, S., McKnight, T., Shah, D. & Meagher, R. (1982) *J. Mol. Appl. Genet.* **1**, 483–498.
11. Schoffl, F., Raschke, E. & Nagao, R. (1984) *EMBO J.* **3**, 2491–2497.
12. Fromm, M., Taylor, L. & Walbot, V. (1986) *Nature (London)* **319**, 791–793.
13. Jefferson, R. A. (1987) *Plant Mol. Biol. Rep.* **5**, 387–405.
14. Murashige, T. & Skoog, F. (1962) *Plant Physiol.* **15**, 473–497.
15. Saiki, R., Gelfand, D., Stoffel, S., Scharf, S., Higuchi, R., Horn, G., Mullis, K. & Erlich, H. (1988) *Science* **239**, 487–494.
16. Esau, K., ed. (1977) in *Anatomy of Seed Plants* (Wiley, New York), pp. 455–570.
17. Rowland, L. J., Chen, Y.-C. & Chourey, P. S. (1989) *Mol. Gen. Genet.* **218**, 33–40.

## Molecular cloning and expression in photosynthetic bacteria of a soybean cDNA coding for phytoene desaturase, an enzyme of the carotenoid biosynthesis pathway

GLENN E. BARTLEY\*, PAUL V. VIITANEN\*, IRIS PECKER†, DANIEL CHAMOVITZ†, JOSEPH HIRSCHBERG†, AND PABLO A. SCOLNIK\*‡

\*Du Pont Central Research and Development, P. O. Box 80402, Wilmington, DE 19880-0402; and †Department of Genetics, The Hebrew University of Jerusalem, Jerusalem 91904, Israel

Communicated by Lawrence Bogorad, April 22, 1991

**ABSTRACT** Carotenoids are orange, yellow, or red photoprotective pigments present in all plastids. The first carotenoid of the pathway is phytoene, a colorless compound that is converted into colored carotenoids through a series of desaturation reactions. Genes coding for carotenoid desaturases have been cloned from microbes but not from plants. We report the cloning of a cDNA for *pds1*, a soybean (*Glycine max*) gene that, based on a complementation assay using the photosynthetic bacterium *Rhodobacter capsulatus*, codes for an enzyme that catalyzes the two desaturation reactions that convert phytoene into  $\zeta$ -carotene, a yellow carotenoid. The 2281-base-pair cDNA clone analyzed contains an open reading frame with the capacity to code for a 572-residue protein of predicted  $M_r$  63,851. Alignment of the deduced *Pds1* peptide sequence with the sequences of fungal and bacterial carotenoid desaturases revealed conservation of several amino acid residues, including a dinucleotide-binding motif that could mediate binding to FAD. The *Pds1* protein is synthesized *in vitro* as a precursor that, upon import into isolated chloroplasts, is processed to a smaller mature form. Hybridization of the *pds1* cDNA to genomic blots indicated that this gene is a member of a low-copy-number gene family. One of these loci was genetically mapped using restriction fragment length polymorphisms between *Glycine max* and *Glycine soja*. We conclude that *pds1* is a nuclear gene encoding a phytoene desaturase enzyme that, as its microbial counterparts, contains sequence motifs characteristic of flavoproteins.

Carotenoids are photoprotective pigments present in all photosynthetic and many nonphotosynthetic organisms. Absorption of visible light and photoprotection are mediated by a chain of conjugated double bonds, the chromophore, that is formed by successive desaturations of the colorless precursor phytoene (1, 2). Deduced peptide sequences of several microbial carotenoid desaturases have been reported (3-6). A maize gene (*yl*) involved in carotenoid biosynthesis has been cloned (7), but at present it is not clear whether this gene is regulatory or structural.

Carotenoids are precursors of abscisic acid (8) and carotenoids influence chloroplast differentiation. Abscisic acid and chloroplast differentiation affect nuclear gene expression (9-11). Thus, a knowledge of the genes of the carotenoid-abscisic acid pathway is required for our understanding of the interactions between plastid differentiation and nuclear gene expression during plant development.

Norflurazon is a bleaching herbicide that inhibits the desaturation of phytoene (12). To isolate a soybean phytoene desaturase gene, we have used a heterologous probe from a gene responsible for norflurazon resistance in the cyanobacterium *Synechococcus* PCC7942 (13). We reasoned that since norflurazon inhibits the desaturation of phytoene, the *Synechococcus*

gene could code for phytoene desaturase and could be used as a heterologous probe for cloning the corresponding plant genes. We report here the cloning and characterization of a soybean cDNA coding for phytoene desaturase.<sup>§</sup>

### MATERIALS AND METHODS

**Recombinant DNA Techniques.** Cloning steps were described in ref. 14. DNA restriction endonucleases, polymerases, and ligases were used under conditions recommended by suppliers (New England Biolabs, Bethesda Research Laboratories, and Perkin-Elmer/Cetus). The PCR was used to amplify a portion of the *Synechococcus* PCC6301 genomic DNA using the primers 5'-TTCTTCGGTGCCTAC-CC(C,T)AA(C,T)AA(C,T)ATG and 5'-GAACAATTTT-TCAAT(T,C)TC(G,A)GCCAT corresponding to the protein sequences FFGAYPNM and TAEIEKLF of the *Synechococcus* PCC7942 protein that confers resistance to norflurazon (13). Hybridization conditions for library screening were as described in Berlyn *et al.* (15). The plasmid pNFPD11 was constructed by ligating a 1857-base-pair (bp) *Sst* I fragment of *pds1* cDNA into the *Sst* I site of pNF3 (16).

**Complementation of *Rhodobacter (Rb.) capsulatus* and Carotenoid Analysis.** Plasmid pNFPD11 was mobilized by conjugation (17) into *Rb. capsulatus* BPY69. Ammonia-free minimal medium was used for induction of the *nif* promoter (16). *Rb. capsulatus* cells were grown photosynthetically in liquid medium and carotenoids were extracted (18) and saponified (19). The solvent used for thin layer chromatography was 5% (vol/vol) acetone in petroleum ether.

**Chloroplast Import.** Linearized plasmid templates were transcribed using T7 RNA polymerase (20). *In vitro* translations were performed using [<sup>35</sup>S]labeled methionine (1100 Ci/mmol; 1 Ci = 37 GBq) and rabbit reticulocyte lysate (Promega) according to the vendor's protocol. Translation reactions were terminated with 2× import buffer (20), containing 60 mM unlabeled methionine (20). Chloroplasts were obtained (21) from 10- to 12-day-old pea seedlings (*Pisum sativum*). Import was as described in ref. 21, but ice instead of HgCl<sub>2</sub> was used to stop reactions. To distinguish between bound and imported polypeptides, plastids were treated with thermolysin (20). Controls were treated identically but received no protease. Intact plastids were repurified by centrifugation through Percoll (22), washed once with import buffer (containing 5 mM EDTA), and then subjected to lysis by resuspending pellets in 250  $\mu$ l of 10 mM Hepes-KOH, pH 8/5 mM EDTA. Total membranes (thylakoids plus envelopes) were recovered by centrifugation at 47,000 × *g* for 30 min. The membranes were washed once and resuspended in gel sample buffer (20). The stromal fraction was recentrifuged

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: nt, nucleotide(s).

†To whom reprint requests should be addressed.

§The sequence reported in this paper has been deposited in the GenBank data base (accession no. M64704).

1 GAATTCCTTCTACGCTACTGCCGTGGTCTTCCACTGCTTACCCTAA  
51 CCTTCTCTCTCTCTGCGCTGCAAGCTTGGTACTCTCAACTCAATTC  
101 TCCACCTTATCTTCTTCTGCTTCTGCTTCTTCTTCTTCTTCTTCTTCT  
151 TTTCAAGTGCCTGAATCTCTGCAACAGTAATTAACACTCTCTCTTCT  
201 GTTCAGGCTTTATTTCCCAATGGCGCTTGGCTATATATCTGCTGCC  
1 MAACGYISAA  
251 AACTTCAATTATCTCGTTGGCGCCAGAACATATCCAAATTCGCTTCTTC  
11 NFNYLVGARNISKFAS  
301 AGACGCCACAATTTCTGCTTCTGCGGGAGCGACTCAATGGGTCTTA  
28 D A T I S F S F G G S D S M G L T  
351 CTTTGGGACCGCTCCGATTCTGCTCTTAAGAGGAACCATTTCTCTCCC  
45 L R P A P I R A P K R N H F S P  
401 TTGCTGTCTGTTTGGCTCGATTATCCAGCCAGAGCTCGAAAAACCGT  
61 L R V V C V D Y P R P E L E N T V  
451 TAATTTCTGTTGAAGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT  
78 N F V E A A Y L S S T F R A S P R  
501 GTCTCTAAACCTTGAACATCGTTATTCGGGTGCGAGATTGGCTGGT  
95 P L K P L N I Y **I** A **E** **A** **G** **L** **A** **G**  
551 TTATCAATGCAAAATTTGGCTGATGCTGGGCATAAACCTATATGCT  
111 L S T A K Y L A D A G H K P I L L  
601 GGAAGCAAGAGAGCTTCTAGTTGGAAAGGTTGCTGCATGGAAGACAAG  
128 **E** **A** **R** **D** **V** **L** **G** **G** **K** **V** **A** **A** **W** **K** **D** **K**  
651 ATGGAGACTGGTACGAGACAGGCTACACATCTTTTGGGGCTTACCCT  
145 G D W Y E T G L H I F F G G A Y A P  
701 TATGTGCGAACCTTTTGGAGAACTGGCATTAAATGATCGGTTACATG  
161 Y V Q N L F G E L G I N D R L Q W  
751 GAAAGAGCATTCTATGATTTTGTCTATGCCAAATAAGCTGGAGATTTA  
178 K E H S M I P A M P N K P G E F S  
801 GTCGATTGATTTTCTGAGTCTTCTCCCTCCCATGGAATGGAATGAG  
195 R F D F P E V L P S P L N G I W  
851 GCAATATTGAGGAACATGAGATGCTTACATGGCCAGAGAAAGTAAAT  
211 A I L R N N E M L T W P E K V K F  
901 TGCAATTTGGGCTTCTCCAGCTATGCTTGGCGGACGCCATATTTGAGG  
228 A I G L L P A M L G G Q P Y V E A  
951 CTCAGATGGTCTTCTGTTCAAGATGGATGAAAAGCAGGCGCTACCT  
245 Q D G L S Y Q E W M K K Q G V P  
1001 GAACGGGTAGCTGATGAGGTTTCAATGCAATGTCAAAGCACTAACTT  
261 E R V A D E V F I A M S K A L N F  
1051 CATCAATCTGATGAATTTCAATGCAATGTATGATTGCTTTAAAC  
278 I N P D E L S M Q C I L I A L N R  
1101 GATTCTTCAGGAGAAATGGTCTAAGATGGCTTTTGGATGGCAAT  
295 F L Q E K H G S K M A F L D G N  
1151 CCACCCGAAAGACTTTGTATGCCAATAGTTGATTATATCAGTCTTGGG  
311 P P E R L C M P I V D Y I Q S L G  
1201 TGGTGAAGTTTCAATTTCCGCGATTCAAAATTTGAGCTAAATGATG  
328 G E V H L N S R I Q K I E L N D D  
1251 ATGGAACGGTGAAGAGTCTTCTAATAATGGGAAGTGAAGGAGG  
345 G T V K S G F L L N N G K V M E G  
1301 GATGCTTATGTTTGGCACTCAGTGGATATCTGAAGCTTCTTCTACC  
361 D A Y Y F A T P V D I L K L L P  
1351 AGATACTGGAAGGATTCATATTTCCAGAGATTGGATAAATAGTTG  
378 D N W **K** G I P Y F Q R **L** D K L V G  
1401 GGGTCCAGTCATAAATGTCATATGTTGACAGAAAGTGAAGAAC  
395 V P V I N V H I W F D R K L K N  
1451 ACATATGATCACTTCTCTTATGAGAGTCCCTTCTGAGTGTATATGC  
411 T Y D **H** L L F S R S P L S V Y A  
1501 TGACATGTCAGTAAGTGAAGGAATATTATAGCCCAACAGTCAATGT  
428 D M S V T T C K E Y **Y** S P N Q S M L  
1551 TAGATTGGTTTTGGTCCAGGAGCAATGATTACGATGATGATGAT  
445 E L V F A P A E E W I S R S D D  
1601 GATATTATCAAGCCAGCTGACTGAGCTTGCCTTCTTCTTCTGATGA  
461 D I I Q A C T M T E L A K L F P D E  
1651 AATTCTGCAGACCAAGCAAGCAAGATCTCAAGTACCATGTTGTTA  
478 I S A A G D Q S K K I L K Y H V V K  
1701 AAACCAAGGTGGTTTACAAAGTGTCCAAATTTGTAACCTTGTGCA  
495 T P R S V Y K T V P N C E P C R  
1751 CCCATTCAAAGATCTCCTATAGAAGTTCTATTAGCTGGAGATTACAC  
511 **E** I Q R S P I E G F Y L A G D Y T  
1801 AAAACAAAAATTTAGCTTCAATGGAAGGCGCTGTTCTTCTGGGAAGC  
528 K Q K Y L A S M E G A V L S G K L  
1851 TTTGTGACAGGCTATTGTACAGGATTCTGAGCTACTAGCTACGGGGC  
545 C A Q A I V Q D S E L L A T R G  
1901 CAGAAAAGATGGCTAAGCAAGTGTGTGAACAAAAACAAGATTGAA  
561 Q K R M A K A S V V  
1951 AGAGTCATGGTACAGTACAGGAGCATCTTCAACTTTGGCATTCTTTGT  
2001 CTGTGGTCCAGACTCAGGAGACTTCACTTTATTAGTTCAACGAATAA  
2051 AGAAAGGCTCAGCTTCTGAAATTTAGCTGCACCGCTGCTCACTGTGCA  
2101 ATAAGCTATACGGAACAAACGACATGTGCACTTTAAGTCAGGCCATTG  
2151 TTTTGTATCTTCAATTTCTGGATCAATGTTTGTATTGGAAGAAATA  
2201 TGTCATTATCAACTGTGTTATATCCACTTTTTTATTATCAACATT  
2251 GTCACAACTTTCTGTTGAGTAAAAAAGAAATTC

FIG. 1. Nucleotide sequence of the *pds1* cDNA coding strand and deduced amino acid sequence of the phytoene desaturase protein. Nucleotides are numbered in the 5' to 3' direction. The coding region starts at nt 221. *Eco*RI sites (GAATTC) flanking the *pds1* cDNA are italicized. Amino acid residues with a match score of at least 3 in the phytoene desaturase pattern (see Fig. 2) are underlined. Residues showing perfect conservation (match score = 5) are shown in bold-

at  $140,000 \times g$  for 30 min and the resultant supernatant was removed for SDS/PAGE. This second centrifugation step greatly improved the analysis of "stromal" carotenoid desaturase, presumably by the partial removal of ribulose 1,5-bisphosphate carboxylase. Samples were subjected to SDS/PAGE on 10% gels. Gel sample solubilization, electrophoresis, and fluorography were carried out as described (20).

## RESULTS

**cDNA Cloning, Nucleotide Sequence, and Predicted Protein Sequence.** A radiolabeled *Synechococcus* PCC6301 probe was used to screen  $5 \times 10^5$  plaques of a soybean immature cotyledon bacteriophage  $\lambda$  cDNA library. Two positive clones isolated showed inserts of similar size (data not shown). A plasmid (pSoysk) derived from one  $\lambda$  clone was purified, and both strands were sequenced by the dideoxynucleotide termination method (23). The 2281-bp insert contains a single open reading frame that stops at nucleotide (nt) 1931, a TAA codon (Fig. 1). The first initiation codon is located at nt 221. Thus, if this ATG is the initiation codon, the cDNA has 200 bp of 5' untranslated sequence, a 1710-bp coding region, and 322 bp of untranslated 3' sequence, followed by a poly(A) tail.

The assignment of the ATG codon at nt 221 as the start codon of *pds1* is supported by three lines of evidence. (i) This codon occurs within the sequence CCAATGGC, which approximates the plant ribosome-binding consensus sequence (25). (ii) As shown below, this cDNA encodes a precursor protein (prPds1) that is active in an *in vitro* chloroplast import reaction, suggesting that all *pds1* 5' sequences coding for the prPds1 N-terminal region required for import are present in this cDNA. (iii) Based on primary sequence conservation (see below), we calculate that the prPds1 transit peptide has about 94 amino acid residues. If the ATG at nt 221 is not the start codon, the transit peptide would exceed 170 residues, which seems unlikely (26).

The *pds1* open reading frame encodes a 572-amino acid protein, with an estimated molecular weight of 63,851. The 322-bp untranslated 3' region contains two putative poly(A) addition signals at nt 2046 (AATAAA) and nt 2100 (AATAA). However, these sequences are located further upstream of the polyadenylation site than the usual 15–40 nt found in many plant genes (27).

**Primary Sequence Conservation Among Prokaryotic and Eukaryotic Carotenoid Desaturases.** To identify amino acid residues that, by virtue of being conserved among different carotenoid desaturases, could be important in structure and function (28), we applied a local alignment algorithm (ref. 29; Fig. 2A) to the deduced peptide sequences of four carotenoid desaturases, *Rb. capsulatus* CrtI and CrtD, *Neurospora crassa* Al-1, and soybean Pds1. The resulting pattern for the family contains 501 characters with 17 conserved residues (Fig. 2B). The alignment "tree" (Fig. 2C) shows that CrtI and Al-1 are the most closely related sequences (match score = 215.27), followed by CrtD (match score = 142.29) and Pds1 (match score = 80.13). Pds1 amino acid residues conserved with a match score of 3 or higher are shown in Fig. 1. From the N terminus the first region of conservation has the sequence Gly-Xaa-Gly-Xaa<sub>2</sub>-Gly-Xaa<sub>3</sub>-Ala-Xaa<sub>4</sub>-Gly. This is characteristic of dinucleotide-binding folds that mediate binding of FAD or NAD(P) (31). This motif has been observed (5, 24) at the N-terminal region of microbial carotenoid desaturases. Extensive conservation has been reported between carotenoid desaturase residues that mediate FAD binding in glutathione reductase and dihydroliipoamide de-

face type. Putative FAD-binding residues (24) are boxed. Amino acid residues are shown in the standard one-letter code.

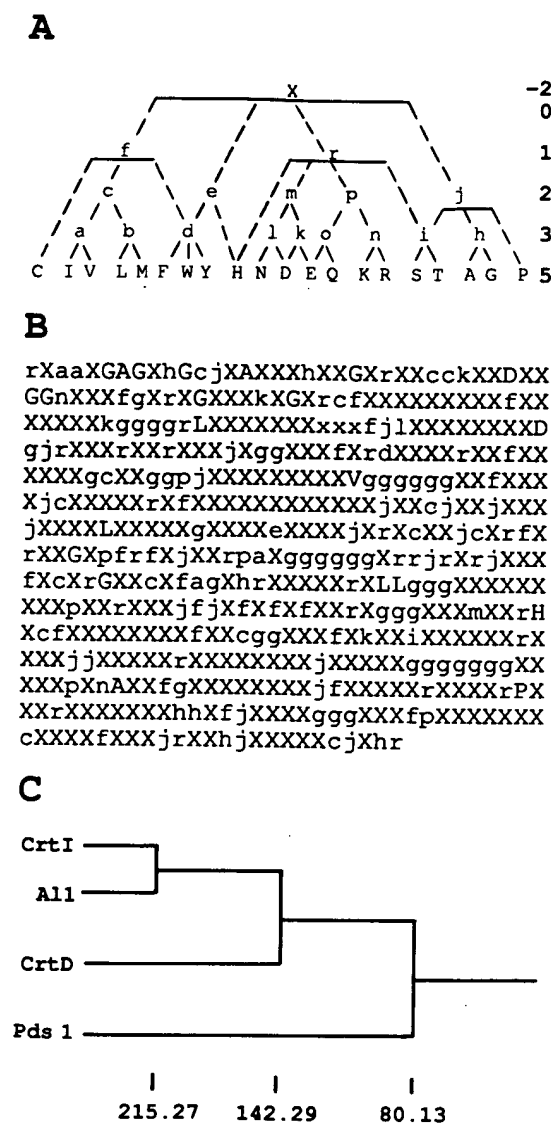


FIG. 2. Primary protein sequence pattern from microbial and plant carotenoid desaturases. (A) Amino acid class hierarchy used to construct the pattern. Uppercase characters indicate standard one-letter code; lowercase characters, amino acid classes; X, any amino acid. Match scores are indicated on the right (modified from ref. 29). (B) Carotenoid desaturase similarity pattern, based on *Rb. capsulatus* CrtI (3) and CrtD (30), *N. crassa* Al-1 (4), and soybean Pds1. g, Gap character. (C) Dendrogram generated, including overall matching scores. For this alignment we changed CrtD at position 13 from R to G assuming a *crtD223* mutation at the corresponding codon (24).

hydrogenase, two disulfide oxidoreductases (24). Based on similar comparison of Pds1 to glutathione reductase and dihydrolipoamide dehydrogenase, we identify 19 Pds1 residues that could play a role in FAD binding (Fig. 1).

**In Vitro Import of Pds1 into Pea Chloroplasts.** In all microbial carotenoid desaturases analyzed, the dinucleotide binding fold is located near the N terminus (5, 24). This plus the facts that these proteins are encoded by nuclear genes and are targeted to plastid envelopes (2, 32) suggest that the soybean Pds1 protein is initially synthesized as a higher molecular weight precursor, possessing a transit peptide that targets the protein to plastids. To test this we conducted *in vitro* chloroplast import experiments. Translation of a *pds1* transcript resulted in the synthesis of a polypeptide (Fig. 3, lane T) whose apparent molecular mass in SDS/PAGE gels ( $\approx 59$  kDa) is in reasonable agreement with the size of the precursor predicted from the DNA sequence.

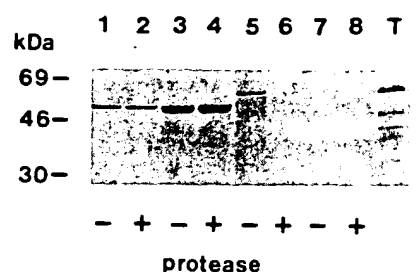


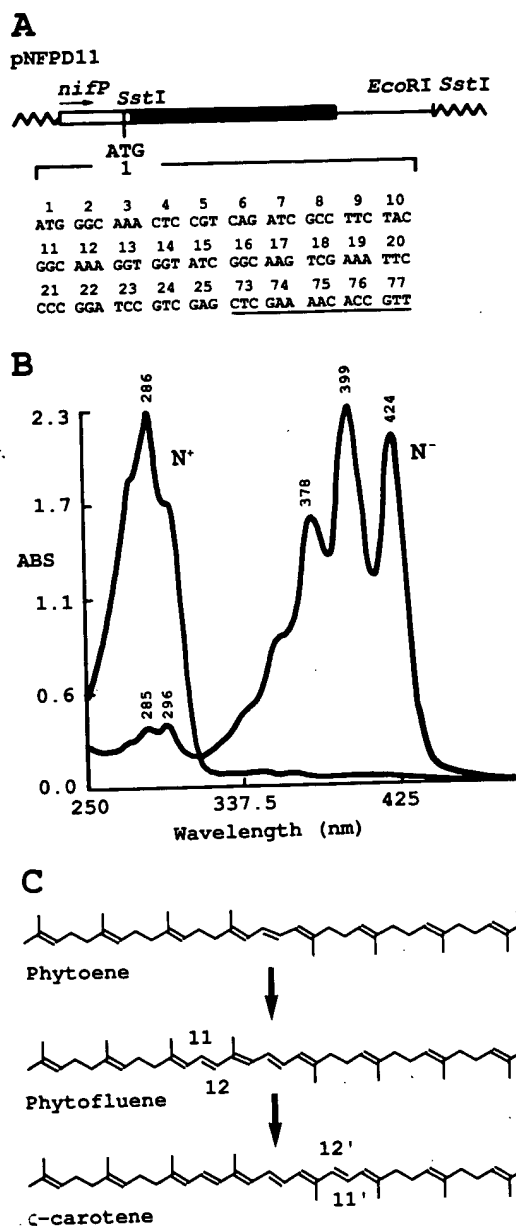
FIG. 3. Import of Pds1 into chloroplasts. Import reactions were carried out in the presence (lanes 1–4) or absence (lanes 5–8) of 5 mM ATP. The latter reactions also contained nigericin ( $0.26 \mu\text{M}$ ) and valinomycin ( $0.88 \mu\text{M}$ ) to completely inhibit the production of ATP by photophosphorylation (33). After import, the plastids with (+) or without (–) protease treatment were fractionated to yield total membranes (lanes 1, 2, 5, and 6) and membrane-free stroma (lanes 3, 4, 7, and 8). A photograph of the SDS/PAGE fluorogram is shown. All lanes received an equivalent amount of chloroplasts. Lane T contains translation products used for import.

The radiolabeled polypeptide was then used for *in vitro* import into isolated pea chloroplasts. After import, intact chloroplasts were recovered and treated with protease, and membrane and soluble plastid fractions were isolated and analyzed by SDS/PAGE (Fig. 3). When import experiments were conducted in the presence of ATP (lanes 1–4),  $^{35}\text{S}$ -labeled prPds1 entered the chloroplasts (as judged by protease resistance) and was processed to a smaller mature form. Surprisingly, imported mature Pds1 was detected in both soluble and membrane fractions. The former could represent either stromal material or soluble material liberated from the envelope intermembrane space. Although the mobility in SDS/PAGE is similar for soluble and membrane-associated Pds1, at present we do not know whether any differences exist in primary sequence between these forms. Regardless, no protease-resistant polypeptides were observed in any chloroplast fraction when import experiments were conducted in the complete absence of ATP (Fig. 3, lanes 5–8). Thus, like other precursor polypeptides targeted to chloroplasts, the import of prPds1 is energy dependent (33).

**Complementation of a Phytoene-Accumulating *Rb. capsulatus* Mutant by *pds1*.** To determine whether the cloned cDNA indeed codes for a phytoene desaturase, we constructed a plasmid that incorporates the *pds1* region coding for the mature protein under the control of bacterial transcription and translation signals. The resulting plasmid pNFPD11 (Fig. 4) contains the promoter, ribosome-binding sequence, and first 25 codons of *Rb. capsulatus nifH* fused in-frame to *pds1* codon 73. The *nif* (nitrogen fixation) promoter can be induced by depriving cells of a fixed nitrogen source (16). Plasmid pNFPD11 was mobilized by conjugation into *Rb. capsulatus* BPY69, a strain with a genetic lesion in *crtI*, the gene that codes for phytoene desaturase (18). When BPY69(pNFPD11) cells were grown in minimal medium without fixed nitrogen, cultures turned yellow and, when cells were grown in ammonia-supplemented medium, they remained blue-green, the color of BPY69 cells lacking the soybean cDNA clone (data not shown). Pigments were extracted and analyzed by absorption spectroscopy (Fig. 4B). The yellow pigment in BPY69(pNFPD11) cells grown in ammonia-free medium was identified as 15-*cis*- $\zeta$ -carotene, indicating that Pds1 mediates two successive desaturations from phytoene (Fig. 4D). The identity of the compound was confirmed by TLC analysis, where it was shown to comigrate with tomato fruit  $\zeta$ -carotene (data not shown).

Based on its structure and function, we assign this gene the name *pds1* (phytoene desaturase 1).

**Genetic Mapping of *pds1*.** The *pds1* cDNA was hybridized to blots containing *Glycine max* and *Glycine soja* DNA digested with 12 restriction enzymes (data not shown). The

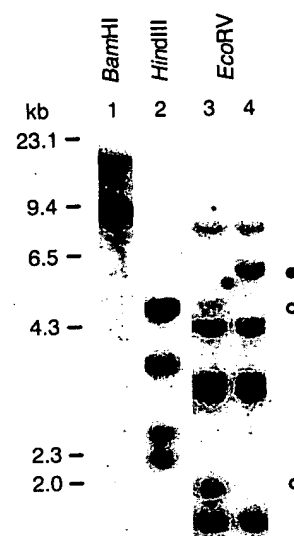


**FIG. 4.** Complementation of a *Rb. capsulatus* carotenoid mutant by the *pds1* cDNA. (A) Construction of the *nif* expression plasmid pNFDP11. Open rectangle, *Rb. capsulatus nif* sequences up to *nifH* codon 25 and including the *nif* promoter (*nifP*) and translation initiation signal of pNF3 (16); solid boxes, *pds1* coding sequences fused to *nif* DNA at the *Sst*I site (GAGCTC, codon 73). The first five *pds1* codons are underlined. Wavy lines represent vector DNA. The *Eco*RI site is located at the 3' end of the *pds1* cDNA (see Fig. 1). (B) Absorption spectra in petroleum ether of carotenoids accumulated by *Rb. capsulatus* BPY69 (pNFDP11) grown in minimal medium with ( $N^+$ ) or without ( $N^-$ ) ammonia. Absorption maxima are indicated for each peak. Phytoene and  $\zeta$ -carotene accumulate under  $N^+$  and  $N^-$  conditions, respectively. (C) Desaturations mediated by Pds1. Carbon atoms at which desaturation takes place are numbered.

patterns obtained with *Eco*RI, *Bam*HI, and *Hind*III suggest that *pds1* is a member of a low-copy-number nuclear gene family (Fig. 5). Polymorphisms detected with *Eco*RV (Fig. 5) allowed mapping of one allele to linkage group 9 in a soybean restriction fragment length polymorphism map (A. Rafalski and S. Tingey, personal communication).

### DISCUSSION

We have cloned a soybean cDNA encoding phytoene desaturase, an enzyme of the carotenoid biosynthesis pathway.



**FIG. 5.** Genomic blots and genetic mapping of *pds1*. The enzymes indicated above the lanes were used to digest DNA from *G. soja* (lanes 1-3) or *G. max* (lane 4). Solid and open circles indicate *G. max* and *G. soja pds1* alleles, respectively, detected with *Eco*RV.

Results indicate that Pds1 is a nuclear-encoded chloroplast protein that catalyzes two successive desaturation reactions.

The *pds1* cDNA was isolated from a cDNA library prepared from immature soybean cotyledons. The low number of positive clones (2 in  $5 \times 10^5$  clones) suggest that this is a low-abundance transcript. Indeed, we failed to detect the *pds1* transcript in Northern blots containing up to 30  $\mu$ g of total soybean RNA or 2  $\mu$ g of poly(A)<sup>+</sup> RNA (data not shown). Although at present we do not know whether the cDNA isolated is a full-length clone, the complementation and chloroplast import results indicate that all of the information necessary for enzymatic activity and plastid targeting is present in the cDNA and suggest that the N terminus of the cloned protein is most likely intact.

Upon import into chloroplasts, mature Pds1 was detected in the soluble and membrane fractions. Carotenoids are membrane pigments and, based on immunological (3) and biochemical (32, 34) analyses, phytoene desaturase and all subsequent enzymes in the pathway are also membrane-bound. Thus, it seems logical to assume that the enzymatically active form of Pds1 is also associated with membranes. This raises questions as to the origin of the soluble Pds1 protein noted above. It is possible that this form is a nonphysiological artifact that accumulates only during the artificial conditions used during *in vitro* chloroplast import. Alternatively, soluble Pds1 may be an intermediate in the membrane insertion process or it may play a nonenzymatic role in either the chloroplast stroma or envelope intermembrane space.

We have not yet defined the chloroplast import processing site. From a number of experiments similar to that shown in Fig. 3, it can be estimated that upon import the precursor polypeptide is shortened by 5-7 kDa. However, estimates obtained from denaturing gels are frequently associated with rather large errors. For microbial carotenoid desaturases, the dinucleotide-binding fold is located 4-9 amino acid residues downstream from the N terminus (4, 5, 24). Thus, if the situation with Pds1 is similar, the mature form would have



≈478 residues and a deduced molecular mass of 54 kDa. This would be in the same range as *N. crassa* and *Rb. capsulatus* phytoene desaturases (24) and is in general agreement with the observed mobility of mature Pds1 in SDS/PAGE gels (Fig. 3). This line of reasoning suggests that the precursor of Pds1 possesses an N-terminal transit peptide of ≈9.8 kDa.

We identified several amino acid residues that are conserved among Pds1, CrtI, CrtD, and Al-1. Among these conserved residues, there is an N-terminal region with the characteristic motif of a dinucleotide binding fold (Fig. 1). However, several amino acid residues that are conserved in all microbial desaturases analyzed are not conserved in Pds1. Notably absent in Pds1 is the C-terminal region containing the His-Pro dipeptide, which is proposed to be equivalent to the His-Pro active site motif of flavoprotein oxidoreductases (24). Thus, the presence of this dipeptide motif is not necessary for the enzymatic conversion of phytoene into  $\zeta$ -carotene.

Expression in an *Rb. capsulatus* carotenoid mutant of the portion of *pds1* coding for the mature protein resulted in the synthesis of  $\zeta$ -carotene. The most likely explanation for this result is that the function of Pds1 is to catalyze the two desaturations needed to convert phytoene into  $\zeta$ -carotene. Alternatively, the specificity of Pds1 may be changed in bacteria. However, both genetic and biochemical evidence indicates that plants have a desaturase that converts phytoene into  $\zeta$ -carotene. Mutant plants of several species blocked in carotenoid desaturations accumulate either phytoene or  $\zeta$ -carotene, suggesting that two distinct genes code for two desaturases (for review, see ref. 1). Beyer *et al.* (34) showed that the conversion of phytoene into 15-*cis*- $\zeta$ -carotene and the subsequent conversion of this pigment into lycopene are two distinct enzymatic reactions. Hence, it appears that plants have two carotenoid desaturases and the simplest explanation for our results is that we have cloned a cDNA for the first of the two enzymes.

We thank Du Pont personnel for the following contributions: K. Abremski (protein alignments), K. Bacot (chloroplast import), A. Rafalski (mapping *pds1*), K. Ripp (cDNA library), J. Williams (design of PCR oligonucleotides), and M. Orbach and S. Russell (critical reading of the manuscript). We also thank P. Beyer (Freiburg) for his assistance in the identification of *cis*- $\zeta$ -carotene.

1. Bramley, P. M. & Mackenzie, A. (1988) in *Current Topics in Cellular Regulation*, eds. Horecker, B. L. & Stadtman, E. R. (Academic, New York), Vol. 29, pp. 291–332.
2. Bartley, G. E., Coomber, S. A., Bartholomew, D. & Scolnik, P. A. (1991) in *Cell Culture and Somatic Cell Genetics of Plants*, eds. Bogorad, L. & Vasil, I. (Academic, New York), Vol. 7B, pp. 331–345.
3. Bartley, G. E. & Scolnik, P. A. (1989) *J. Biol. Chem.* **264**, 13109–13113.
4. Schmidhauser, T. J., Lauter, F. R., Russo, V. E. A. & Yanofsky, C. (1990) *Mol. Cell. Biol.* **10**, 5064–5070.
5. Armstrong, G. A., Alberti, M. & Hearst, J. E. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 9975–9979.
6. Misawa, N., Nakagawa, M., Kobayashi, K., Yamano, S., Izawa, Y., Nakamura, K. & Harashima, K. (1990) *J. Bacteriol.* **172**, 6704–6712.
7. Buckner, B., Kelson, T. L. & Robertson, D. S. (1990) *Plant Cell* **2**, 867–876.
8. Creelman, R. A. (1989) *Physiol. Plant.* **75**, 131–136.
9. Giuliano, G. & Scolnik, P. (1988) *Plant Physiol.* **86**, 7–9.
10. Burgess, D. G. & Taylor, W. C. (1988) *Mol. Gen. Genet.* **214**, 89–96.
11. Bartholomew, D. M., Bartley, G. E. & Scolnik, P. A. (1991) *Plant Physiol.* **96**, 291–296.
12. Clarke, I. E., Bramley, P. M., Sandmann, G. & Böger, P. (1982) in *Biochemistry and Metabolism of Plant Lipids*, eds. Wintermans, J. F. & Kuiper, P. J. C. (Elsevier, Amsterdam), pp. 549–554.
13. Chamovitz, D., Pecker, I. & Hirschberg, J. (1991) *Plant Mol. Biol.* **16**, 967–974.
14. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
15. Berlyn, M. B., Last, R. L. & Fink, G. R. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 4604–4608.
16. Pollock, D., Bauer, C. E. & Scolnik, P. A. (1988) *Gene* **65**, 269–275.
17. Giuliano, G., Pollock, D., Stapp, H. & Scolnik, P. A. (1988) *Mol. Gen. Genet.* **213**, 78–83.
18. Giuliano, G., Pollock, D. & Scolnik, P. A. (1986) *J. Biol. Chem.* **261**, 12925–12929.
19. Davies, B. H. (1976) in *Chemistry and Biochemistry of Plant Pigments*, ed. Goodwin, T. W. (Academic, New York), Vol. 2, pp. 65–66.
20. Viitanen, P. V., Doran, E. R. & Dunsmuir, P. (1988) *J. Biol. Chem.* **263**, 15000–15007.
21. Reed, J. E., Cline, K., Stephens, L. C., Bacot, K. O. & Viitanen, P. V. (1990) *Eur. J. Biochem.* **194**, 33–42.
22. Cline, K., Werner-Washburne, M., Lubben, T. H. & Keegstra, K. (1985) *J. Biol. Chem.* **260**, 3691–3696.
23. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
24. Bartley, G. E., Schmidhauser, T. J., Yanofsky, C. & Scolnik, P. A. (1990) *J. Biol. Chem.* **265**, 16020–16024.
25. Lütcke, H. A., Chow, K. C., Mickel, F. S., Moss, K. A., Kern, H. F. & Scheele, G. A. (1987) *EMBO J.* **6**, 43–48.
26. Keegstra, K. & Olsen, L. J. (1989) *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **40**, 471–501.
27. Mogen, B. D., MacDonald, M. H., Graybosch, R. & Hunt, A. G. (1990) *Plant Cell* **2**, 1261–1272.
28. Berg, P. & Reichardt, J. K. V. (1988) *Nucleic Acids Res.* **16**, 9017–9026.
29. Smith, R. F. & Smith, T. F. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 118–122.
30. Armstrong, G. A., Alberti, M., Leach, F. & Hearst, J. E. (1989) *Mol. Gen. Genet.* **216**, 254–268.
31. Moller, W. & Amons, R. (1985) *FEBS Lett.* **186**, 1–7.
32. Lütke-Brinkhaus, F., Liedvogel, B., Kreuz, K. & Kleinig, H. (1982) *Planta* **156**, 176–180.
33. Grossman, A., Bartlett, S. & Chua, N.-H. (1980) *Nature (London)* **285**, 625–628.
34. Beyer, P., Mayer, M. & Kleinig, H. (1989) *Eur. J. Biochem.* **148**, 141–150.



# The Encoded Primary Sequence of a Rice Seed ADP-glucose Pyrophosphorylase Subunit and Its Homology to the Bacterial Enzyme\*

Claim 67

ADP-Glucose  
pyrophosphorylase

(Received for publication, March 6, 1989)

Joseph M. Anderson†, James Hnilo‡, Raymond Larson‡§, Thomas W. Okita‡, Matthew Morell¶||, and Jack Preiss¶

From the ‡Institute of Biological Chemistry, Washington State University, Pullman, Washington 99164-6340 and the  
¶Department of Biochemistry, Michigan State University, East Lansing, Michigan 48824

Rice seed ADP-glucose pyrophosphorylase cDNA clones were isolated by screening a  $\lambda$  expression library prepared from rice endosperm poly(A<sup>+</sup>) RNA with a heterologous antibody raised against the spinach leaf enzyme and subsequently by nucleic acid hybridization. One cDNA plasmid, possessing about 1650 nucleotides, was shown by both DNA and RNA sequence analysis to contain the complete ADP-glucose pyrophosphorylase coding sequence of 483 amino acids. The primary sequence displayed a putative leader peptide presumably required for transport of this nuclear encoded protein into the amyloplasts, a differentiated starch containing plastid. The leader peptide, however, showed little sequence homology with transit peptides displayed by other known nuclear encoded proteins localized in the chloroplasts. A comparison of the primary sequence of the putative mature subunit to the *Escherichia coli* pyrophosphorylase showed two regions displaying significant homology. These two conserved regions contain residues shown previously to be essential for the allosteric regulation and catalytic activity of the *E. coli* enzyme. Differences in the primary sequences of the plant and bacterial enzyme may reflect the distinct nature of the allosteric effectors that control these enzymes.

ADP-glucose pyrophosphorylase plays a pivotal role in the control of synthesis of starch in plants (Preiss *et al.*, 1967; Preiss, 1988) and of glycogen (Preiss, 1984) in bacteria. This enzyme catalyzes the synthesis of the activated glucosyl donor, ADP-glucose, from Glc-1-P and ATP. The catalytic activity of both the plant and bacterial ADP-glucose pyrophosphorylases is strongly affected by small effector molecules. Although the nature of these effector molecules differ

among bacterial and plant enzymes, they typically are major metabolites which accumulate during normal carbon metabolism in these organisms (Preiss, 1978). The *Escherichia coli* enzyme is activated by Fru-1,6-P<sub>2</sub> and inhibited by ADP and AMP, whereas the plant enzyme is stimulated and suppressed by 3-PGA<sup>1</sup> and orthophosphate, respectively (Preiss, 1978). Results obtained from both intact leaf and chloroplast systems support the concept that starch biosynthesis is controlled by the ratio of 3-PGA to P<sub>i</sub> via the allosteric control of ADP-glucose pyrophosphorylase *in vivo* (Heldt *et al.*, 1977). Allosteric regulation of ADP-glucose pyrophosphorylase from maize endosperm (Plaxton and Preiss, 1987) and potato tuber (Sowokinos and Preiss, 1982) has also been demonstrated *in vitro*, although it is not clear whether the levels of these effector molecules oscillate in these developing nonphotosynthetic tissues, as detected in chloroplasts of leaf tissue. Nevertheless, the activity of the above two nonphotosynthetic ADP-glucose pyrophosphorylases is highly dependent on the presence of the allosteric activator, 3-phosphoglycerate.

ADP-glucose pyrophosphorylases from both *E. coli* and *Salmonella typhimurium* are homotetramers with subunit molecular weights of about 49,000 (Preiss, 1984). Unlike the bacterial enzymes, the plant ADP-glucose pyrophosphorylase exhibits more structural complexity. The homogeneous spinach leaf enzyme was observed to have a native molecular weight of 206,000 (Copeland and Preiss, 1981) and composed of dissimilar subunits of 54,000 and 51,000 by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (Copeland and Preiss, 1981; Morell *et al.*, 1987a). These subunits can be distinguished on the basis of their N-terminal sequences, tryptic peptide patterns, and immunological reactions (Morell *et al.*, 1987a). Immunological related subunits of similar sizes have also been observed in leaf tissue of *Arabidopsis thaliana* (Lin *et al.*, 1988a, 1988b), wheat, rice, and maize (Krishnan *et al.*, 1986). The isolation of two nuclear mutations that affect the level of one or both subunits in *Arabidopsis* (Lin *et al.*, 1988a, 1988b) supports the postulate that the plant enzyme is a product of two different genes.

Functional studies with pyridoxal phosphate, an analog of the allosteric activator 3-PGA, suggest a role for both subunits in enzyme function (Morell *et al.*, 1987b, 1988). To learn more about the structure of the plant enzyme we have cloned the complete coding sequence of an ADP-glucose pyrophosphorylase transcript from rice endosperm. Structural analysis has shown that this plant subunit shares significant homology at the amino acid level with the *E. coli* polypeptide. Despite the relative conservation of sequences near the N terminus, shown

\* This work was supported in part by Grant DE-FG06-87ER1 3699 from the United States Department of Energy, The Rockefeller Foundation, Project 0590, College of Agriculture and Home Economics Research Center, Washington State University (to T. W. O.), and National Science Foundation Grant DMB86-10319 (to J. P.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) J04960.

§Present address: Dept. of Veterinary Microbiology and Pathology, Washington State University, Pullman, WA 99164.

||Present address: Plant Environmental Biology Group, Research School of Biological Sciences, Australian National University, Canberra, A.C.T. 2601 Australia.

<sup>1</sup> The abbreviations used are: 3-PGA, 3-phosphoglycerate; ORF, open reading frame; bp, base pair(s).

earlier to be essential for allosteric regulation of the bacterial enzyme (Larsen *et al.*, 1986; Parsons and Preiss, 1978), residues essential for catalytic control of the plant enzyme are located near the C terminus (Morell *et al.*, 1987b, 1988). As discussed in more detail herein, we suggest that the differences in the allosteric specificity exhibited by the bacterial and plant enzyme may be accounted for, at least in part, by structural differences mediated by specific segmental and point mutations.

#### MATERIALS AND METHODS

**Plant Material**—Rice seeds (*Oryza sativa* L. cv. Biggs M201) were germinated and grown in an environment-controlled growth chamber as described earlier (Krishnan *et al.*, 1986). Seeds were collected at 12–15 days after flowering and stored at  $-80^{\circ}\text{C}$ .

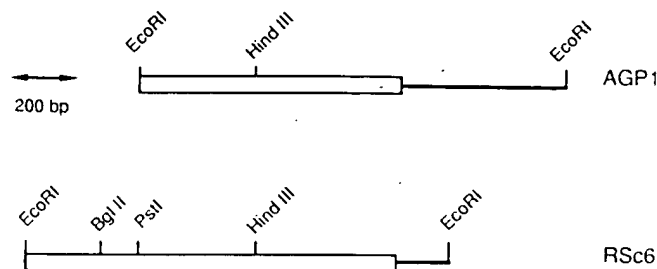
**cDNA Library Construction and Screening**—All labware and solutions, where appropriate, were sterilized by baking at  $180^{\circ}\text{C}$  for at least 2 h or by autoclaving to eliminate ribonucleases. Total RNA was isolated from 12–15 days after flowering seeds as described previously (Krishnan *et al.*, 1986). Polyadenylated (poly(A<sup>+</sup>)) RNA was selected by oligo(dT)-cellulose chromatography as described by Maniatis *et al.* (1982). A cDNA library was prepared in  $\lambda$  gt11 and screened with either antiserum, prepared against spinach leaf holoenzyme of ADP-glucose pyrophosphorylase (Krishnan *et al.*, 1986), or  $^{32}\text{P}$ -labeled AGP-1 cDNA obtained by nick translation (Maniatis *et al.*, 1982).

**DNA Sequencing and Primer Extension**—The cDNA clones were subcloned into M13-mp18, and sequential deletions were prepared using exonuclease III as described by Henikoff (1984). Single-stranded DNAs from an ordered set of deletions were isolated and sequenced by the Sanger dideoxy chain termination method (Sanger *et al.*, 1977). Sequence analyses were performed with the UWGCC program (Devereux *et al.*, 1984). Direct RNA sequencing was carried out by reverse transcribing poly(A<sup>+</sup>) RNA in the presence of a 16-mer synthetic oligonucleotide complementary to the region 20 bases downstream from the 5'-end of the RSc6 cDNA clone. The primer extension reaction was performed essentially as described by Inoue and Cech (1985).

#### RESULTS AND DISCUSSION

**Cloning of ADP-glucose Pyrophosphorylase mRNA**—In a previous study, a cDNA clone, AGP-1, was isolated by screening a  $\lambda$  gt11 library with antibodies raised against the spinach leaf ADP-glucose pyrophosphorylase (Krishnan *et al.*, 1986). The structure of this cDNA clone was determined by DNA sequence analysis to be a cloning artifact. Although the 5'-end segment of the DNA insert contained an open reading frame (ORF) of 1035 bp for ADP-glucose pyrophosphorylase, it displayed an unusually long 3'-segment of over 670 bp and lacked a poly(A) tail (Fig. 1). As an additional 500–600 nucleotides would be required to complete the remaining coding sequence and poly(A) tract, the total length of this cDNA would greatly exceed the estimated transcript size of 1800 nucleotides determined by Northern blot analysis (Krishnan *et al.*, 1986). The 3'-noncoding sequence of this clone was observed to be unreactive to rice seed poly(A<sup>+</sup>) RNA or DNA indicating that it was a spurious contaminant isolated during the cloning procedure. To obtain an authentic recombinant clone, a second library was constructed in  $\lambda$  gt11 and screened by nucleic acid hybridization using the AGP-1 insert as a probe. Fourteen clones were isolated. One of these clones, pRSc6, contained a cDNA insert of 1650 nucleotides and was chosen for further study.

**Sequence and Amino Acid Analysis**—The 1650-bp cDNA clone RSc6 was sequenced on both strands by dideoxyribonucleotide chain termination reactions using overlapping deleted clones as templates (Fig. 1). An ORF of 483 residues was found which begins with an ATG at bp position 1 and ends at the termination codon TAA at bp position 1460 (Fig. 2). This ORF of 483 amino acids predicts a protein with a



**FIG. 1. Physical map of restriction enzyme sites for endosperm-specific ADP-glucose pyrophosphorylase cDNAs.** The cDNAs were purified from the  $\lambda$  gt11 cloning vector and inserted into Bluescript plasmid DNA. The resulting plasmid DNA was digested with restriction enzymes that digested the cloning vector only once. AGP-1 contained DNA sequences that encoded a truncated subunit possessing 341 residues from the C terminus. The 3'-end of AGP-1 contained 28 nucleotides of the 3'-untranslated region and an additional 646 bp of non-pyrophosphorylase sequence inadvertently cloned during the construction of the cDNA library. A second cDNA clone, RSc6, was isolated from a second cDNA library using  $^{32}\text{P}$ -labeled AGP-1 DNA as a probe. Open rectangles represent the coding sequences for ADP-glucose pyrophosphorylase.

molecular weight of approximately 52,000 consistent with the size of the polypeptide of 56,000 estimated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis of *in vitro*-translated product (Krishnan *et al.*, 1986). The termination codon is followed by a 3'-untranslated region of 201 bp and a poly(A) tail. The consensus polyadenylation signal, AAUAAA, typical of most nonhistone eukaryotic mRNAs (Proudfoot and Brownlee, 1976), is present but is 117 bp upstream from the polyadenylation addition site. A variant form, AAUGAU, is observed within the usual 10–30 bp of the poly(A) site and may serve as the signal.

As the cDNA clone began with an ATG, it was not clear whether these nucleotides represented the initiation codon or an internal methionine residue. Therefore, RNA sequencing studies were initiated. A synthetic 16-base oligonucleotide, complementary to nucleotides 20–35 of the messenger RNA transcript, was prepared, annealed to poly(A<sup>+</sup>)-RNA, and subjected to chain termination reactions using reverse transcriptase. An additional 27 nucleotides at the 5'-end were obtained by this method and revealed two nonsense codons. One of these termination codons lies 15 bases upstream of the putative ATG start and is in-frame with the coding sequence (Fig. 2). The presence of a nonsense codon upstream of the putative translational start indicates that the ORF present in pRSc6 is representative of the complete ADP-glucose pyrophosphorylase coding sequence. The initiator codon is located in the sequence CAACCATCA which is consistent with the consensus translational start sites of both animal (Kozak, 1984) and plant (Lutcke *et al.*, 1987) transcripts.

ADP-glucose pyrophosphorylase is located exclusively in the chloroplasts of leaves (Okita *et al.*, 1979) and the amyloplast in developing endosperm and potato tuber tissue.<sup>2</sup> As this enzyme is encoded by nuclear genes, it was expected that the protein would possess a leader sequence required for plastid localization. The N-terminal sequence of the mature subunit has not been elucidated, and, therefore, the exact processing site cannot be specifically assigned. However, alignment of the rice primary sequence to the N-terminal sequence of the purified spinach leaf 51,000 pyrophosphorylase subunit obtained by Edman degradation (Morell *et al.*, 1987b, 1988) indicates partial homology and suggests that the proteolytic cleavage site occurs between residues 28 and 29

<sup>2</sup>W. T. Kim, V. R. Francheschi, T. W. Okita, N. L. Robinson, M. Morell, and J. Preiss, manuscript in preparation.

TTTGTAAATCTTTAATTTGTTGCAACC<sup>-1</sup>

ATGAAATGTGTTGGCATCTAAGATCTTCCCTTCCCGCTCCAATGTTGTTAGCGAGCAACAACATCGAAGCGC<sup>72</sup>  
 M N V L A S K I F P S R S N V V S E Q Q Q S K R<sup>24</sup>  
 GAGAAAGCAACTATTGATGACGCTAAGAACTCGTCCAAGAACAAAAATCTTGACCGCAGTGTGATGAGAGT<sup>144</sup>  
 E K A T I D D A K N S S K N K N L D R S V D E S<sup>48</sup>  
 GTGCTTGGAAATCATCTCTGGAGGTGGTGCAGGGACTAGATTGTATCCCTCACCAGAGCGTGCCAAAGCCT<sup>216</sup>  
 V L G I I L G G G A G T R L Y P L T K K R A K P<sup>72</sup>  
 GCCGTGCCTCTGGGTGCCAACTACAGGCTTATTGATATCCCTGTGACCAACTGTTTGAACAGCAACATATCC<sup>288</sup>  
 A V P L G A N Y R L I D I P V S N C L N S N I S<sup>96</sup>  
 AAGATCTATGTGCTGACACAATTCAACTCTGCCTCTCCGAACCGTCACTGTCAAGAGCCTATGGGAACAAC<sup>360</sup>  
 K I Y V L T Q F N S A S P N R H L S R A Y G N N<sup>120</sup>  
 ATTGGCGGGTACAAGATGAAGGGTTCGTTGAAGTCTCGCTGCACAGCAGAGCCAGATAATCTTAATCGG<sup>432</sup>  
 I G G Y K N E G F V E V L A A Q Q S P D N P N W<sup>144</sup>  
 TTTCAAGGTACTGCAGATGCTGTAAGACAGTACTTATGGCTATTTGAGGAGCATAATGTTATGGAGTTTCTA<sup>504</sup>  
 F Q G T A D A V R Q Y L W L F E E H N V M E F L<sup>168</sup>  
 ATTCTGGCTGGAGATCACCTTTACCGCATGGACTATGAAAGTTCAITCAGGCACACAGAGAAACAGATTCT<sup>576</sup>  
 I L A G D H L Y R M D Y E K F I Q A H R E T D S<sup>192</sup>  
 GATATTACTGTGCTGCCCTGCCAATGGATGAGAAACGTGCAACTGCATTCGGCTCATGAAAATTGACGAG<sup>648</sup>  
 D I T V A A L P M D E K R A T A F G L M K I D E<sup>216</sup>  
 GAAGGAGAATAGTTGAATTTGCAGAGAAACAAAAGGAGAACAAATTGAAAGCAATGATGGTTGACACTACC<sup>720</sup>  
 E G R I V E F A E K P K G E Q L K A M M V D T T<sup>240</sup>  
 ATACTCGCCTTGACGACGTGAGGGCAAAGGAAATGCCTTATATGTAGCATGGGTATCTATGTGATTAGT<sup>802</sup>  
 I L G L D D V R A K E M P Y I A S M G I Y V I S<sup>264</sup>  
 AAAATGTAAATGCTTCAGCTCCTCCGTGAACAATTTCTGGAGCTAATGACTTTGGAAGTGAAGTCAATCCA<sup>874</sup>  
 K N V M L Q L L R E Q F P G A N D F G S E V I P<sup>288</sup>  
 GGTGCAACAAACATCGGAATGAGGGTACAAGCTTACTTGTATGATGGTTACTGGGAAGACATTGGTACCATT<sup>946</sup>  
 G A T N I G M R V Q A Y L Y D G Y W E D I G T I<sup>312</sup>  
 GAGGCATTCTATAATGCAATCTGGGAATAACGAAAAGCCTGTACCAGATTTCAGTTTCTATGACCGGCT<sup>1018</sup>  
 E A F Y N A N L G I T K K P V P D F S F Y D R S<sup>336</sup>  
 GCTCCAAATTTATACACAACCTCGACACTTGCTCCTCTCGAAGGTTCTTGATGCTGATGTGACAGACAGTGC<sup>1090</sup>  
 A P I Y T Q P R H L P P S K V L D A D V T D S V<sup>360</sup>  
 ATTGGTGAAGGGTGTGTTATTAATAAAGTGTAGATACACCATTCAGTAGTTGGACTCCGGTCTGCATTCT<sup>1162</sup>  
 I G E G C V I K N C K I H S V V G L R S C I S<sup>384</sup>  
 GAAGGCGCAATAATAGAGGACTCATTACTCATGGGAGCTGACTACTACGAGACTGAAGCAGACAAGAACTC<sup>1234</sup>  
 E G A I I E D S L L M G A D Y Y E T E A D K K L<sup>408</sup>  
 CTTGGTGAAAAGGTGGCATTCCCATTTGGTATTTGGGAAGAATTGCCACATTAGAAGGGCAATCATCGACAAG<sup>1306</sup>  
 L G E K G G I P I G I G K N C H I R R A I I D K<sup>432</sup>  
 AATGCTCGTATTGGAGATAATGTGAAGATAATCAATGTTGACAAATGTCCAAGAAGCTGCAAGAGAGACTGAT<sup>1378</sup>  
 N A R I G D N V K I I N V D N V Q E A A R E T D<sup>456</sup>  
 GGATACTTCATCAAAAGTGGCAITGTTACCGTGATCAAGGATGCTTTGCTCCTAGCGGAACAGTTATATGAA<sup>1452</sup>  
 G Y F I K S I V T V I K D A L L L A E Q L Y E<sup>480</sup>  
 GTAGCTCGTAATATATGATGGGGCATCGGCGACGAGCACCAGGCGGGCATTATAAGAAGAAATAAAGCAGT<sup>1522</sup>  
 V A A \*<sup>483</sup>  
 TCAAGGCTCTTTATCCCTTTTTTTTCTTCTGTTCTTCTCTATGTTTTAGGATAAATGTAATGGTAGAA<sup>1594</sup>  
 GCAACTGCTTTTCAGATGTTTTGGGAGTGAATGATACCGTGCTTGAATTTTCC<sup>1647</sup>

Fig. 2. DNA and deduced primary sequence of RSc6. The DNA sequence was obtained by analysis of recombinant clones containing overlapping segments of DNA as described in the text. Numbers on the right are in bp or amino acid residues relative to the translational start. The 5'-untranslated sequences were obtained by RNA sequencing of rice seed poly(A<sup>+</sup>) RNA as described in the text. The translation initiation and polyadenylation signals are underlined.

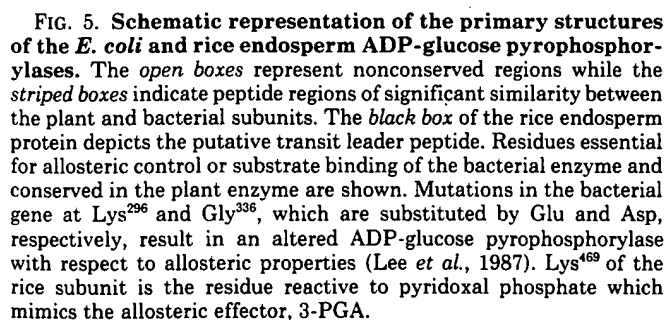
Rice MNV L A S K I F P S R S N V V S E Q Q Q S K R E K A T I D D A K N S S K N K N L D R S  
 Spinach V S D S Q N S Q D G - - L D P E

Fig. 3. Alignment of the N-terminal regions of the encoded rice polypeptide to the mature spinach leaf *M*, 51,000 pyrophosphorylase subunit determined by Edman degradation (Morell *et al.*, 1987a). Residues conserved between the rice and spinach sequences are enclosed in boxes. Gaps were introduced to maximize the homology. The arrow designates the putative proteolytic cleavage site for the rice-encoded polypeptide. The hydrophobic N terminus and charged region enriched in basic residues of the rice pyrophosphorylase leader sequence are overlined.

(Fig. 3). The putative leader sequence of ADP-glucose pyrophosphorylase, however, shares very little sequence conservation particularly with the three blocks of homology suggested by Karlin-Neumann and Tobin (1986) to be necessary for chloroplast transport and processing. The pyrophosphorylase leader sequence does contain short peptide domains

possessing physical properties similar to those observed in other chloroplast transit peptides. The N terminus possesses a 6-residue hydrophobic domain and a highly basic region near the C-terminal end of the leader peptide (Fig. 3), and these features may serve as the signals necessary for plastid targeting and processing of this plant protein.

FIG. 4. Conserved primary sequences displayed between the rice endosperm (*R*) and *E. coli* (*E*) ADP-glucose pyrophosphorylases. Residues conserved among these two proteins are enclosed in boxes. Gaps, indicated by dashes, were introduced to maintain homology. Numbers labeled on the right side of the figure indicate the number of residues from the translational start of either the rice or bacterial subunit. Lys<sup>39</sup>, Tyr<sup>114</sup>, Lys<sup>195</sup>, Lys<sup>296</sup>, and Gly<sup>336</sup> residues essential for allosteric and/or catalytic function of the bacterial enzyme are underlined. The conserved rice sequence at residues 462–472 homologous to the pyridoxylated peptide and which presumably lies in the allosteric region of the spinach leaf *M*<sub>1</sub>51,000 subunit is overlined.



creased to 42%. Nonconserved regions of the plant and bacterial subunits were evident at the N termini encompassing residues 29–47, a central domain between residues 228 and 296, and at the C termini beginning with residue 395. The significant conservation displayed in the peptide domain spanning amino acids 48–227 is particularly noteworthy as specific residues have been identified that are essential for allosteric regulation and substrate binding of the bacterial enzyme (Parsons and Preiss, 1978). Lysine residues at positions 39 and 195 of the *E. coli* primary sequence have been shown to be protected from phosphopyridoxylation by the activator, Fru-1,6-P<sub>2</sub>, and the substrate ADP-glucose (Baecker *et al.*, 1983; Parsons and Preiss, 1978), suggesting that these 2 residues lie at or near the allosteric and substrate binding sites. Likewise, the tyrosine residue at position 114 in the *E. coli* enzyme when altered by site-directed mutagenesis to phenylalanine causes lower affinity for its activator, fructose 1,6-P<sub>2</sub>, and substrates, ATP and ADP-glucose, and altered a lower apparent binding for its inhibitor, AMP (Kumar *et al.*, 1988; Larsen *et al.*, 1986). The conservation of other amino acids in this region in the plant enzyme is strongly suggestive that these amino acids may also contribute a major role in maintenance of protein conformation and/or the regulatory and catalytic functions of the plant enzyme with 3-PGA as the major activator and P<sub>i</sub> the inhibitor.

As the bacterial and plant enzymes catalyze the same reaction the structural differences exhibited by the plant and bacterial enzymes most likely reflect the nature of the effector molecules that mediate allosteric regulation of each enzyme. Other than the low homology exhibited between residues 228 and 296, major structural changes are evident at the C termini of the primary sequences between the plant and bacterial enzymes (Figs. 4 and 5). Morell *et al.* (1987b, 1988) have shown that pyridoxal phosphate, which mimics the activator 3-PGA, reacts upon reduction to a specific lysine residue of the spinach leaf 51,000 subunit. The reactive residue, however,

was not the conserved Lys<sup>68</sup> whose counterpart, Lys<sup>39</sup>, lies within the allosteric region of the *E. coli* subunit but within the sequence Ser-Gly-Ile-Val-Thr-Val-Ile-Lys-Asp-Ala-Leu. This peptide, which is observed in the rice endosperm subunit, is located at the nonconserved C terminus of the plant enzyme (Figs. 2 and 4) suggesting that at least a portion of the primary structure responsible for regulation of the plant enzyme lies on an extended peptide not evident in the bacterial subunit. In addition the change in allosteric specificity of the plant enzyme is most likely due to amino acid replacements in other regions of the plant enzyme. Lee *et al.* (1987) and Kumar *et al.* (1989) have shown that point mutations of the bacterial gene where Lys<sup>296</sup> and Gly<sup>336</sup> are substituted by glutamic and aspartic acid residues, respectively, result in the formation of a bacterial enzyme whose activity is less dependent on allosteric activation by Fru-1,6-P<sub>2</sub>. Interestingly, residues of the plant subunit which directly align with Lys<sup>296</sup> and Gly<sup>336</sup> of the bacterial subunit are negatively charged amino acids (Figs. 4 and 5). Point mutations to acidic residues at positions 329 and 363 and addition of a C-terminal peptide to the plant enzyme may have simultaneously resulted in the loss of allosteric specificity to Fru-1,6-P<sub>2</sub> but a gain in sensitivity to 3-PGA. Whether these point mutations and structural differences are a manifestation of the nature of the effector molecules remains to be explored. The allosteric and catalytic behavior of a recombinant enzyme constructed by the translocation of the plant C-terminal peptide to the bacterial allosteric mutant enzyme may provide such information.

## REFERENCES

- Baecker, P. A., Furlong, C. E., and Preiss, J. (1983) *J. Biol. Chem.* **258**, 5084-5088
- Copeland, L., and Preiss, J. (1981) *Plant Physiol.* **68**, 996-1001
- Devereux, J., Haeberli, P., and Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387-395
- Heldt, H. W., Chon, J., Maronde, D., Stankovic, Z. S., Walker, D. A., Kraminer, A., Kirk, M. R., and Heber, V. (1977) *Plant Physiol.* **59**, 1146-1155
- Henikoff, S. (1984) *Gene (Amst.)* **11**, 4891-4906
- Inoue, T., and Cech, T. R. (1985) *Proc. Natl. Acad. Sci. U. S. A.* **82**, 648-652
- Karlin-Neumann, G. A., and Tobin, E. M. (1986) *EMBO J.* **5**, 9-13
- Kozak, M. (1984) *Nucleic Acids Res.* **12**, 857-872
- Krishnan, H., Reeves, C. D., and Okita, T. W. (1986) *Plant Physiol.* **81**, 642-645
- Kumar, A., Tanaka, T., Lee, Y. M., and Preiss, J. (1988) *J. Biol. Chem.* **263**, 14634-14639
- Kumar, A., Ghosh, P., Hill, M. A., and Preiss, J. (1989) *J. Biol. Chem.* **264**, 10464-10471
- Larsen, C. E., Lee, Y. M., and Preiss, J. (1986) *J. Biol. Chem.* **261**, 15402-15409
- Lee, Y. M., Kumar, A., and Preiss, J. (1987) *Nucleic Acids Res.* **15**, 10603
- Lin, T-P., Caspar, T., Somerville, C., and Preiss, J. (1988a) *Plant Physiol.* **86**, 1131-1135
- Lin, T-P., Caspar, T., Somerville, C., and Preiss, J. (1988b) *Plant Physiol.* **88**, 1175-1181
- Lutcke, H. A., Chow, K. C., Mickel, F. S., Moss, K. A., Kern, H. F., and Scheele, G. A. (1987) *EMBO J.* **6**, 43-48
- Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Morell, M. K., Bloom, M., Knowles, V., and Preiss, J. (1987a) *Plant Physiol.* **85**, 182-187
- Morell, M., Bloom, M., Larsen, C., Okita, T. K., and Preiss, J. (1987b) *UCLA Symp. Mol. Cell. Biol. New Ser.* **62**, 227-242
- Morell, M., Bloom, M., and Preiss, J. (1988) *J. Biol. Chem.* **263**, 633-637
- Okita, T. W., Greenberg, E., Kuhn, D. N., and Preiss, J. (1979) *Plant Physiol.* **83**, 187-192
- Parsons, T. F., and Preiss, J. (1978) *J. Biol. Chem.* **253**, 7638-7645
- Plaxton, W. C., and Preiss, J. (1987) *Plant Physiol.* **83**, 105-112
- Preiss, J. (1978) *Adv. Enzymol. Relat. Areas Mol. Biol.* **46**, 317-381
- Preiss, J. (1984) *Annu. Rev. Microbiol.* **38**, 419-458
- Preiss, J. (1988) in *The Biochemistry of Plants* (Preiss, J., ed) Vol. 14, pp. 184-249, Academic Press, Orlando, FL
- Preiss, J., Ghosh, H. P., and Wittkop, J. (1967) in *Biochemistry of Chloroplasts* (Goodwin, T. W., ed) Vol. 2, pp. 131-153, Academic Press, New York
- Proudfoot, N. J., and Brownlee, C. C. (1976) *Nature* **263**, 211-214
- Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74**, 5463-5467
- Sowokinos, J. R., and Preiss, J. (1982) *Plant Physiol.* **69**, 1459-1466

### **VIII. Argument**

#### **A. REJECTION OF CLAIMS OVER THE GOLDMAN ET AL. '073 PATENT**

##### ***SUMMARY OF THE REJECTION***

Each of the pending claims has been finally rejected on the basis of both obviousness and anticipation over the Goldman et al. '073 patent. The position taken in the final Action is that (1) Goldman is at least enabling for what it claims (a process for introducing DNA into corn, without any representation that the process enables the preparation of fertile corn); and (2) that Goldman positively shows that "hygromycin resistance" and "EPSP synthase" activity had been achieved using *Agrobacterium*-mediated transformation.

In a rejection being considered together with the foregoing, the Action posits that Appellants' specification admits that the particular genes of the pending claims are known in the prior art. The Action reasons that at least certain of these references disclose the introduction of the gene(s) into other plant species, and this provides the requisite motivation for introduction of these genes into corn.

##### ***APPELLANTS' LEGAL BASIS AND SUMMARY OF ARGUMENT***

With respect to the anticipation aspect of the rejection, Appellants note that none of the presently pending claims recites subject matter that is in any way taught or disclosed in Goldman. The independent claims currently on appeal are directed to fertile, transgenic maize plants that have introduced one of more transgenes selected from a *bar* gene, a nitrilase gene, a  $\beta$ -lactamase gene, an

xylE gene, a galactosidase gene, a tyrosinase gene, or a wheat germ agglutinin gene. Appellants have been unable to identify any disclosure in Goldman related to any of these particular individual genes. Thus, it is respectfully submitted that any concern regarding anticipation is misplaced.

With respect to the obviousness concerns, in order to maintain the rejection, the Examiner must demonstrate that a case of obviousness of the appealed claims is established by the teachings of the '073 patent specification considered in combination with the secondary references. *In re Deuel*, 34 U.S.P.Q.2d 1210, 1214 (Fed. Cir. 1995). The claims of the present appeal require fertile, transgenic corn plants bearing one or more of the specified genes. To be relevant to patentability, the '073 patent, alone or in combination with secondary teachings, must provide:

- (1) detailed enabling methodology for reproducibly preparing fertile, transgenic corn bearing the individual genes specified in the claims;
- (2) a suggestion to introduce the particular genes into corn; and
- (3) evidence suggesting that the particular genes would function appropriately upon successful introduction into corn, and that functional genes would be transmitted progeny plants.

*O'Farrell*, 7 U.S.P.Q.2d at 1680. Goldman, alone or in combination with the secondary references, fails to satisfy the foregoing requisites of *O'Farrell*.

With respect to enablement, the art must do more than simply teach a means of introducing DNA into corn, it must teach a reproducible and reliable method of obtaining fertile, transgenic corn bearing the particular genes specified in the claims. As noted early on by the Supreme Court in the case of *Seymour v. Osborne*, 78 U.S. 516, 20 L. Ed. 33 (1871):

Mere vague and general representations will not support [the invalidity of a later invention], as the publication must be sufficient to enable those skilled in the art or science to understand the nature and operation of the invention, and to carry it into practical use.

78 U.S. at 555.

The evidence of record demonstrates that the method for introducing DNA into corn as set forth in the '073 patent was incapable of enabling the production of fertile, transgenic corn at the time of its priority date, 1987 or 1986. The '073 specification was found to be unreliable for this purpose by both the PTO and an inventor of the '073 patent. As noted by Judge Learned Hand, no doctrine is better established in patent law than that in order for a prior patent to invalidate a later claim, the prior patent must bear adequate directions within its four corners, its results must be reproducible, and it must be "more than a starting point for further experiments":

No doctrine of the patent law is better established than that a prior patent or other publication to be an anticipation must bear within its four corners adequate directions for the practice of the patent invalidated. If the earlier disclosure offers no more than a starting point for further experiments, if its teaching will sometimes succeed and sometimes fail if it does not inform the art without more how to practice the new invention, it has not correspondingly enriched the store of common knowledge and it is not an anticipation.



*Dewey & Almy Chem. Co. v. Mimex Co.*, 124 F.2d 986, 980, 52 U.S.P.Q. 138 (2d Cir. 1942). Thus, to be considered for the purposes of obviousness, a reference must either be fully enabling in and of itself, or be combined with a secondary reference whose teachings will remedy the inadequacies of the first reference. *Symbol Technologies, Inc. v. Opticon, Inc.*, 19 U.S.P.Q.2d 1241 (Fed. Cir. 1991).

In order to demonstrate that a reference in non-enabling, Appellants are required to do so by only a preponderance of the evidence. *In re Sasse*, 207 U.S.P.Q. 107, 111 (CCPA 1980). Once an appellant demonstrates non-enablement by a preponderance of the evidence, the obviousness rejection cannot be maintained unless the Examiner comes forward with evidence negating the appellant's basis of the Goldman patent's non-enablement.

In the present case, Appellants will demonstrate that a preponderance of the evidence compels the conclusion that the '073 patent does not reproducibly and reliably enable the preparation of fertile, transgenic corn, for the reasons that:

- (1) the only evidence of gene transmission to a second generation presented by Goldman employed an improper assay that was not accepted by the PTO in the Goldman prosecution;
- (2) Goldman acquiesced in the earlier finding by the PTO that the '073 specification does not enable fertile, transgenic corn;
- (3) the PTO itself has argued successfully that *Agrobacterium*

was not enabled for the preparation of transgenic monocots at a time well after the Goldman et al. priority date(s) in a published Federal Circuit opinion;

- (4) the particular *Agrobacterium* strains taught by Goldman are now known to be ineffective in the preparation of fertile, transgenic corn; and
- (5) one of the Goldman inventors, Ann Graves, confirms that the technique disclosed in the '073 specification is unpredictable and unreliable.

The foregoing evidence provides a strong indication of inoperability. In response to these indications, the Examiner fails to counter with a scientifically reasonable explanation of why the art would expect the Goldman method to yield fertile, transgenic corn, much less other evidence of operability of Goldman's method.

The second requirement of *O'Farrell* is that the Examiner must demonstrate a suggestion in the art to modify the principal reference to arrive at the subject matter of the claims. Here, the claims relate to fertile, transgenic corn plants that have been genetically engineered to introduce particular gene species. The Examiner must thus show the secondary references provide a suggestion to modify the Goldman teachings with respect to the individual specified genes.

The only secondary references even generally identified by the Examiner are those articles cited in Appellants' specification, that merely detail the isolation and/or *in vitro* manipulation of

the various genes that Appellants have introduced into transgenic corn. However, none of these articles in any way teaches or suggests introducing any of the recited genes into the corn genome. It is indeed Appellant's specification that teaches the introduction of the recited genes into corn and it is improper to use the Appellant's own teachings to support an obviousness rejection. Furthermore, the secondary art fails to provide any discussion relevant to how one would proceed to introduce the genes into corn, fails to provide any evidence that such an introduction would be successful if attempted, and fails to provide any evidence that the genes would function appropriately if introduced into corn.

Lastly, *O'Farrell* requires that there be some evidence that the prior art technique would be successful in the preparation of transgenic corn bearing the particular cited genes, and that the genes would be expected to function appropriately in progeny. If the evidence is insufficient to establish that Goldman had actually achieved fertile, transgenic corn, there is certainly no evidence that the Goldman technique would be successful in the case of the particular genes of the present invention that are not even mentioned in Goldman. Furthermore, because of the unreliability of the opine assays used by Goldman to purportedly demonstrate successful introduction of a gene into corn, there is no evidence that the method disclosed in the Goldman specification would be successful in the case of any gene.

**THE GOLDMAN TECHNIQUE DOES NOT ENABLE THE  
PREPARATION OF TRANSGENIC MAIZE OR PROGENY**

**Agrobacterium Mediated DNA Transfer in Dicotyledonous Plants**

*Agrobacterium tumefaciens* is a bacterium that is used extensively for the introduction of recombinant DNA into dicotyledonous plants (dicots). Wounding of a dicot plant cell results in induction of a series of DNA processing and transfer events in *Agrobacterium* that culminate in the introduction of DNA into the cell by means of the T-DNA (Transferred DNA) portion of the Ti plasmid.<sup>1</sup> The T-DNA integrates into the nuclear genome of the infected plant cell, and the plant cell in turn seeks to express the genes carried on the T-DNA, and indeed manufactures certain enzymes encoded by the T-DNA segment that elicits the synthesis of tumor-specific compounds called opines. Thus, introduction and expression of foreign genes into dicots is achieved through *Agrobacterium* infection by inserting the desired gene into the T-DNA region of the Ti plasmid, followed by appropriate wounding and infection of the dicot. Numerous factors affect the ability of any particular plant species to be susceptible to *Agrobacterium*-mediated gene transfer, particularly whether the species has an appropriate wound response mechanism, and how the particular strain of *Agrobacterium* responds to the wounded cell.

---

<sup>1</sup> A Ti plasmid is an extrachromosomal closed circular DNA that carries genes that will induce crown gall disease in plants, and is used as the DNA carrying vector in *Agrobacterium*-mediated plant transformation.

### **The *Agrobacterium*-based Technique Disclosed by Goldman**

The *Agrobacterium* technique disclosed by Goldman is no different from the technique as it was traditionally practiced in dicots. Goldman suggests making a wound in a *Gramineae* seedling in an area containing rapidly dividing cells, and inoculating the wound with vir+ *Agrobacterium tumefaciens* containing a genetically engineered T-DNA (Exhibit B; col. 6, l. 45-69). After the seedlings are inoculated with the vir+ *A. tumefaciens*, they are incubated until transformation takes place, at which time the seedlings are planted and allowed to grow until they produce pollen. (col. 7, l. 55-60).

### **Why Maize is Not An Appropriate Host for *Agrobacterium***

Maize is not a natural host of *Agrobacterium*, because it lacks at least one key feature required for successful transformation by *Agrobacterium*. Cereals in general lack the wound response that is necessary for successful integrative *Agrobacterium* transformation. Instead, as explained by Potrykus (Exhibit D), cereal cells die when wounded, whereas most dicots and some monocots respond to wounding by production of dedifferentiation and cell proliferation in adjacent cells. Furthermore, most dicot cells are known to produce phenolic compounds following wounding, whereas these compounds are generally not produced by wounded cereal cells. These two features of cereal wound responses, i.e., lack of cell proliferation at the wound site and lack of production of phenolic compounds, are impediments to successful *Agrobacterium*

transformation. Binns (Exhibit E) explained that successful *Agrobacterium* transformation requires that the cells into which *Agrobacterium* T-DNA is transferred must be capable of cell division and monocots typically do not exhibit this response:

Thus, the ability of wounded plant cells to enter and carry out one or more cell cycles may be absolutely required for successful [*Agrobacterium*] transformation. It is interesting to note that the wound response of most monocotyledons is quite different from that of dicotyledon: cells around the wound site lignify or sclerify in the absence of apparent cell division.

(Exhibit E, p. 137, right column). Therefore, a wounded monocot cell is not competent for *Agrobacterium* transformation and the *Agrobacterium* infected cell is fated to die.

Phenolic compounds produced by wounded plant cells are detected by *Agrobacterium* virulence (*vir*) genes, ultimately resulting in the induction of T-DNA processing and DNA delivery to the host cell. Usami et al. (Exhibit F) explained that whereas dicot plants produce diffusible compounds, e.g., phenolic compounds, that are capable of inducing *vir* gene expression, monocot plants, including corn, do not produce these compounds and therefore the process of T-DNA transfer to the host cell is never initiated following *Agrobacterium* infection of a wounded monocot cell. Usami et al. (Exhibit F) further explains that the process of T-DNA transfer can be initiated in a monocot following external application of a known *vir* gene inducing compound (e.g. acetyosyringone). The '073 specification does not teach a method of inducing *vir* gene expression in *Agrobacterium* and, therefore, it is highly unlikely that T-DNA processing and delivery to the host

cell would occur following the methods disclosed in the '073 specification.

**Agroinfection of Maize via *Agrobacterium* Does Not Lead to Stable Transformation and Integration**

While it is true that viral DNA, e.g., Maize Streak Virus (MSV), can be delivered to maize via *Agrobacterium tumefaciens* in a process termed "agroinfection" (Boulton et al., Exhibit G), successful virus infection is based upon the ability of the virus to move from cell to cell and spread to sites away from the wound inoculation site, i.e., away from those cells fated to die in response to wounding. If the DNA were integrated into the host genome of the wounded recipient cell as purported by Goldman, a successful viral infection would not occur because the virus would be unable to move to the healthy proliferating cells away from the wound inoculation site.

**Goldman Teaches the Wrong Strain of *Agrobacterium***

The inoperability of the Goldman disclosure in the preparation of fertile, transgenic corn is further evidenced by the fact that Goldman discloses inoculation of maize seedlings with both octopine and nopaline synthesizing strains of *Agrobacterium* and purport to achieve equal efficiencies of DNA delivery with both types of strains. However, even agroinfection of maize with MSV using *Agrobacterium* as a vector is highly dependent upon the strain of bacteria used. In particular, Boulton et al. demonstrated that whereas it was possible to achieve agroinfection of maize plants

with MSV using nopaline producing strains of *Agrobacterium* such as C58, it was not possible to do so using octopine producing strains such as Ach5 and others (see Table 1, Boulton et al., Exhibit G). Goldman, however, purports to produce fertile, transformed plants in Example XI using *Agrobacterium* strain C19, a derivative of strain Ach5! Thus, it is highly improbable that Goldman could have achieved even non-integrative agroinfection with Ach5-derived strains.

The data in column 16 of the Goldman patent was also produced using an octopine producing strain of *Agrobacterium* which based on Agroinfection data (Boulton et al., Exhibit G) is not capable of delivering DNA to maize cells. The strain of *Agrobacterium* which is capable of delivering DNA to the maize cell is not related to the type of DNA that is being delivered (e.g., viral or non-viral), but is related to the process that *Agrobacterium* uses for delivering any DNA sequence to a plant cell. Therefore, a strain of *Agrobacterium* that does not deliver DNA to maize cells in the process of Agroinfection, can hardly be expected to deliver DNA to maize cells using the process described by Goldman.

**The Assays Employed by Goldman Are Notoriously Unreliable and the Controls Employed Were Inappropriate**

Of particular importance is the fact that Goldman relies only upon the detection of octopine to verify that transformation has occurred. The assays employed by Goldman detect the conversion of arginine to an opine compound, and are notoriously unreliable as proof of the presence of T-DNA. Christou et al. (Exhibit H)



demonstrated that normal callus and plant tissue of several species of plants, including maize (page 220) are capable of converting arginine to nopaline in the absence of T-DNA! Thus, the presence of enzyme activities that will convert arginine to opine compounds is to be expected in maize cells that do not contain T-DNA and hence there is no definitive evidence that the plants disclosed by Goldman are transformed at all. This was the basis of the PTO's § 112(1) rejection in the Goldman prosecution.

The Example in column 16 of the Goldman specification does not purport to disclose fertile transgenic maize plants and progeny, but at best purports to disclose transient expression of introduced DNA, i.e., lysopine dehydrogenase activity was assayed in inoculated seedlings only at 7-14 days after inoculation. Appellants submit that the proper controls were not done in these experiments. Goldman disclosed assays of seedlings that have been infected with strains of *Agrobacterium* that contain functional or non-functional *vir* genes. Both the *vir*<sup>+</sup> and the *vir*<sup>-</sup> strains of *Agrobacterium* used by Goldman contain a functional lysopine dehydrogenase gene. The proper control for this experiment should have been a strain of *Agrobacterium* completely lacking the Ti plasmid, including the lysopine dehydrogenase gene, or uninfected maize seedlings.

Furthermore, the number of samples assayed in the experiment disclosed in column 16 of the Goldman patent is too small to generate data from which conclusions may be drawn. In Figure 3B of the Goldman patent (described in column 16), five maize seedlings

inoculated with a *vir+* strain of *Agrobacterium* were assayed and one seedling was observed to produce a compound that migrated similar to octopine. However, only four *vir-* inoculated seedlings were assayed for lysopine dehydrogenase activity. This is simply not enough seedling assays to conclude that lysopine dehydrogenase activity is not present in *vir-* *Agrobacterium* infected maize seedlings. Goldman et al. did not even assay as many *vir-* *Agrobacterium* inoculated seedlings as they did *vir+* *Agrobacterium* inoculated seedlings! If only four *vir+* *Agrobacterium* inoculated maize seedlings had been assayed, it is possible that Goldman et al. would have concluded that *vir+* *Agrobacterium* strains are not capable of transforming maize!

The importance of sufficient numbers of proper negative controls is underscored by Christou et al. (Exhibit H), who specifically addressed the unreliability of opine assays. Christou et al. concluded that,

"Our results indicate that the presence of opine synthase activity in callus or plant tissue should only be considered a preliminary indication of transformation by *Agrobacterium*. Further biochemical analyses need to be carried out in order to establish whether transformation has taken place. It is not clear if this endogenous synthase activity is physiologically important to the plant or if it is a laboratory artifact of some unknown enzyme that has another catalytic role." (emphasis added, page 221, left column).

In his discussion Christou refers to all opine synthesizing enzymes including lysopine dehydrogenase and nopaline dehydrogenase as opine synthase activities, and states that these activities are not conclusive evidence of transformation. Goldman's only evidence for transformation of maize is the presence of opine synthesizing

activities. Goldman has no evidence beyond purported evidence of opine synthesizing activity in *Agrobacterium* infected maize tissues and certainly does not demonstrate the presence of introduced foreign DNA in maize or even the stability of expression of the purported opine synthesizing activities.

**There is No Teaching in Goldman that "Hygromycin Resistance" or "EPSP Synthase" Expression Had Actually Been Attempted or Accomplished, and No Evidence of any Controls**

Nowhere does Goldman indicate that "hygromycin resistance" and "EPSP synthase" activity had been achieved. On the contrary, the excerpts from Goldman relied upon by the Examiner are clearly prophetic, and give no indication that any assays were ever actually carried out.

The final Action states that Goldman teaches that hygromycin resistance "was obtained" and that proper controls were given, referring to column 21-22. This is submitted to be incorrect: The passage beginning at column 22, line 17, simply describes a hygromycin assay *per se*, written in the present tense rather than the past tense of the previous paragraph. There is no indication in this passage that negative or positive controls were or even should be carried out.

The same can be said for Goldman's purported teaching of transgenic corn expressing "EPSP synthase" activity. The entire "Example X" of Goldman, beginning at col. 22, line 35, is written in the present tense, with no actual studies or results of studies being reported. This is true for the so-called "Assays" of same

reported beginning at col. 23, line 44 -- As with the hygromycin "tests", no tests, no test results!

***DR. ANN GRAVES, ONE OF THE INVENTORS OF THE '073 PATENT,  
AGREES THAT THE TECHNIQUE IT DISCLOSES IS NOT RELIABLE OR  
PREDICTABLE***

Dr. Ann Graves, a named inventor of the '073 patent, has confirmed to the present Appellants that the technique disclosed in the '073 specification is unreliable and generally unpredictable. Dr. Christopher Flick, an employee of the instant assignee, contacted Dr. Graves to obtain her opinion about the invention described in the '073 specification. A declaration of Dr. Flick's setting forth the substance of his various conversations with Dr. Graves is attached hereto as Exhibit I, and further includes a detailing of Dr. Flick's discussions with a notable expert in the field of *Agrobacterium*-mediated plant transformation, Dr. Eugene Nester of the University of Washington.

As detailed in paragraphs 5-10 of the Flick declaration, Dr. Graves has indicated that the *Agrobacterium* mediated transformation methods set forth in the '073 patent were unpredictable and unreliable in their results, and that the techniques were only successful on occasion. Dr. Graves suggested that this was because there was a narrow "window" of susceptibility of Graminae seedlings to *Agrobacterium*, and that there was no means known to her to pinpoint this "window". She further indicated that without additional knowledge, *Agrobacterium* mediated transformation of Graminae, including maize, cannot be reliably carried out.

From the foregoing it can only be concluded that the '073 patent is not enabling for the preparation of fertile, transgenic corn. Dr. Flick indicates that he has reviewed the '073 specification and has confirmed that it fails to teach or suggest the problem pinpointed by Dr. Graves that must be overcome in order to provide an enabling method -- that of identifying the requisite "window" for maize transformation with *Agrobacterium*.

Dr. Flick's declaration is further strengthened by reference to statements made by Dr. Eugene Nester, a recognized expert in the transfer of DNA to plants by means of *Agrobacterium*. Dr. Nester indicated that transfer of DNA to monocots by *Agrobacterium* is highly dependent upon the correct selection of the *Agrobacterium* strain, and that octopine producing strains such as was taught by Goldman are incapable of delivering DNA to monocots such as corn. Furthermore, Dr. Nester indicates that at the time of Goldman's filing, there was no knowledge in existence that would have indicated to those of skill in the art which strains could be employed and under what specified conditions. Therefore, the Goldman specification is not only misleading, there were no teachings at that time to correct the deficiency. Lastly, Dr. Nester confirmed that opine assays such as those relied upon by Goldman are notoriously inaccurate in plant tissues, and concludes that the '073 specification discloses no credible evidence for transformation of maize or any other *Graminae* species.

**THE "LAW OF THE CASE" MUST BE RESPECTED: THE PATENT OFFICE HAS ALREADY DETERMINED THAT GOLDMAN IS NOT ENABLING FOR THE PREPARATION OF FERTILE TRANSGENIC CORN -- AND GOLDMAN ACQUIESCED IN THAT DETERMINATION**

Any allegation that Goldman is enabling for the preparation of fertile, transgenic corn is not supported by the record and is in direct contradiction to the PTO's position in the Goldman prosecution. The PTO consistently took the position on the Goldman record, that the Goldman specification is NOT enabling for the preparation of transgenic corn *per se*, and, at best, merely describes one technique for transferring T-DNA into corn. In this regard, Appellants direct the Board's attention to the file history of USSN 06/880,271, a now-abandoned parent of Goldman *et al.*, in which the Goldman applicants sought allowance of claims directed to transformed corn plants *per se*.<sup>2</sup>

In an Office Action dated 5/30/89 (see Exhibit J), the Goldman examiner rejected a claim directed generally to transgenic corn plants under 35 U.S.C. § 112, first paragraph, stating that:

The specification only demonstrates the expression of heterologous genes, namely opine synthases, in seedling tissue or plant parts directly derived from growing the transformed seedling. No demonstration of stable gene integration or sexual transmission of the exogenous gene, other than prophetic examples, is shown. Given the recalcitrance of monocots to Agrobacterium transformation, as discussed by Applicants, and the possibility of transient opine synthase expression by non-integrative genes, as discussed by Hernalsteens *et al.* (see, e.g., page 3040, column 2, second paragraph), undue experimentation would be required by one of ordinary skill in the art to obtain stable gene integration or sexual transmission of the exogenous gene as claimed.

---

<sup>2</sup> Appellants are not aware of any further pending applications in the Goldman series of applications. If the Examiner is aware of any such applications or patents, the Examiner is requested to make them of record.

In response, Goldman et al. failed to present any additional data per se on the record in support of a conclusion that stable integration had been achieved, merely pointing to new matter added in what is now Example XI of the '073 patent. However, this example merely relates to the purported demonstration of opine synthesis, already found by the PTO Examiner to be an insufficient test for stable gene integration.

In the next Office Action (OA of 3/8/90; Exhibit K), the rejection was maintained, with the examiner finding that stable gene integration had not been demonstrated:

Applicants urge that rejection of the claims as being non-enabled for transformed plants obtained by sexual reproduction is improper, given the demonstration of opine synthase expression in cells directly derived from transformed meristematic cells. The Examiner maintains that opine synthase is an insufficient test for stable gene integration (see, e.g., Christou et al., which discloses opine production by untransformed cells). (emphasis supplied)

The application became abandoned as a result of Goldman's failure to respond to the 3/8/90 Office Action. This abandonment by Goldman of claims to transgenic plants is submitted to be an acquiescence, or at least a rebuttable presumption, that the specification is enabling only for the narrow process set forth in the '073 claims, and not enabling for the production of fertile, transgenic corn plants in general.

In the case of *Litton Systems Inc. v. Whirlpool Corporation*, 221 U.S.P.Q. 97 (Fed. Cir. 1984), the Federal Circuit held that failure to argue and appeal a final rejection on the basis of enablement under § 112, operated as an estoppel against the applicant to later assert that the original specification was

enabling. This principle has been adhered to in the more recent case of *Waldemar Ltd., GmbH & Co. v. Osteonics Corp.*, 31 U.S.P.Q.2d 1855 (Fed. Cir. 1994). That is precisely the situation here: The Goldman applicants attempted to claim transgenic plants, were finally rejected on the basis that there was insufficient evidence that transgenic plants had actually been obtained, and failed to appeal or continue to prosecute claims directed to transgenic plants. It is submitted that by analogy to the *Whirlpool* doctrine, it must be concluded that Goldman acquiesced in the holding of non-enablement, and it is inappropriate to now contravene the previous holding of the PTO and independently "resurrect" the subject matter of Goldman's acquiescence.

No additional evidence in support of a conclusion that the PTO was wrong in its initial determination has been presented; there is no reasoned scientific explanation of record why we should now question the PTO's determination in the very issue at hand -- whether Goldman is enabling for the preparation of transgenic maize plants. Absent additional evidence, under the "law of the case" doctrine, the previous finding of the PTO in the Goldman prosecution itself should be followed. This is not unlike the situation in *Ex parte Edwards*, 231 U.S.P.Q. 981 (PTO Bd. Pat. App. Interf. 1986), where the Board observed that it was bound by an earlier determination that the patent in question was entitled to a particular filing date. To contravene the "law of the case" doctrine, it is incumbent upon the Examiner to provide a cogent explanation as to why the earlier determination by the PTO in the



Goldman prosecution was incorrect and should not be followed.

The Examiner's response to the foregoing position is that "each case is decided on its own merits." While Appellants agree with this proposition in general, it is unclear how this in any way supports a conclusion that Goldman is enabling for fertile, transgenic corn. Since the merits of Goldman is precisely what is at issue here, the fact that each case is decided on its own merits supports Appellants' position. The current "of record" PTO position is that Goldman is non-enabling for fertile, transgenic corn. This holding is thus effectively *res judicata* and binding absent additional evidence.

***THE POSITION THAT AGROBACTERIUM-MEDIATED TRANSFORMATION IS INEFFECTIVE IN CORN IS SUPPORTED BY THE CASELAW***

Appellants' position that Goldman is not enabling for the preparation of fertile, transgenic corn is supported by the recent decision by the Federal Circuit in *In re Goodman*, 29 U.S.P.Q.2d 2010 (Fed. Cir. 1993). The *Goodman* case involved the issue of whether *Agrobacterium*-mediated transformation is effective to transform monocots. The application in question in *Goodman* was filed at approximately the same time as that of Goldman (1985 v. 1986). The Federal Circuit, in confirming the decision of the Board of Appeals, held that *Agrobacterium* mediated transformation was ineffective to transform monocots such as corn, relying on numerous scientific sources for its determination.

One of the principal pieces of art relied upon by the Federal Circuit in its determination was an article authored by Goodman in

1987 -- two years after the filing date of the Goodman application and one year after Goldman's filing date -- which underscored the fact that even at that late date there had been no successes with monocot transformation using *Agrobacterium*. The Federal Circuit also relied upon a review authored by Dr. Ingo Potrykus, in which he stated that "it has been widely considered that monocotyledonous plants, including the commercially important crop plants of the Gramineae family, are insensitive to [Ti plasmid transformation] and thus are not candidates for use of this gene transfer system."

29 U.S.P.Q. at 2014.

The Federal Circuit summed up its finding that at as late as 1987, *Agrobacterium*-mediated monocot transformation was "encumbered" with "great uncertainties," stating:

Thus, even the references cited by Goodman to show enablement support the Board's position that great uncertainties encumbered *Agrobacterium*-mediated transformation in monocot plants at the time of filing. Goodman's 1987 article shows that the claimed invention did not overcome those uncertainties . . . Thus, on Goodman's 1985 filing date, the record shows no reliable gene transformation method for use with monocot plants. Each of the methods for monocot plants was fraught with unpredictability. The teachings in the specification do not cure this unpredictability. The record shows that practicing a gene transformation method for all monocot plants, if possible at all in 1985, would have required extensive experimentation that would preclude patentability.

29 U.S.P.Q. 2014-2015.

**GOLDMAN MUST BE HELD TO THE SAME STANDARD OF PROOF OF  
ENABLEMENT AS HAVE THE CLAIMS OF THE PRESENT CASE**

The claims of the present case have been subject to an exceedingly stringent standard of proof of enablement, with the

Examiner continually taking the position that only those transgenic plants bearing transgenes actually shown to function through actual experimental results are enabled. In the present prosecution, only following the demonstration of the successful introduction and expression of many, many different transgenes in corn has the Examiner agreed that Appellants' gene species have been enabled.

Appellants do not agree with the Examiner's stringent approach in this regard, and are of the opinion that once a reasonable number of species have been demonstrated, that Appellants should be entitled to a generic claim. Nevertheless, as noted above, prior art is subject to the same enablement standard as is an applicant's specification. See, *Dewey & Almy Chem. Co. v. Mimex Co. and Seymour v. Osborne*, supra. It is therefore submitted that in the context of the present prosecution, and the present prosecution only, the Examiner must apply the same rigorous standard to Goldman as has been applied against Appellants. The only "transgene" expression purportedly demonstrated by Goldman was lysopine dehydrogenase.

***THE CASELAW SUPPORTS A CONCLUSION THAT THERE IS NO MOTIVATION TO INTRODUCE THE PARTICULAR GENES SPECIFIED IN THE APPEALED CLAIMS INTO MAIZE***

In order to find the appealed claims obvious, the Examiner must demonstrate that the prior art provides a reproducible method of preparing transgenic maize, that the individual genes specified in the claims are enabled, and that there is sufficient motivation for one of skill in the art to introduce the specified genes. It is Appellants' position that even if the Goldman technique is found

to be broadly enabling for the preparation of transgenic offspring corn, it nevertheless fails to obviate the particular genes specified in the claims absent a motivation to introduce the particular gene into corn, and evidence that these genes would function appropriately upon introduction into corn. *In re Vaeck*, 20 U.S.P.Q. 1438 (Fed. Cir. 1991).

The claimed combination of elements in each of the claims -- which specify particular foreign genes introduced into the corn genome -- is a novel combination. As such, it is incumbent upon the Examiner to identify individual teaching(s) that would suggest the desirability of making each of the claimed combinations. As recently pointed out by the Federal Circuit, in assessing the patentability of a novel combination of otherwise old elements, "[t]he critical inquiry is whether there is something in the prior art as a whole to suggest the desirability, and thus the obviousness, of making the combination." *In re Newell*, 13 U.S.P.Q.2d 1248, 1250 (Fed. Cir. 1989).

The requirement that examiner's identify such a motivation is a longstanding patent law doctrine, even in biological cases. In the well-known case of *In re Bergel and Stock*, the CCPA stated it thusly:

The mere fact that it is possible to find two isolated disclosures that might be combined in such a way to produce a new compound does not necessarily render such production obvious unless the art also contains something to suggest the desirability of the proposed combination.

130 U.S.P.Q. 206 (CCPA 1961).

Motivation to combine elements can not be inferred from prior art that enables one of the elements of the combination. This is the clear meaning of the relevant case law, such as *In re Gordon*, wherein the Federal Circuit stated:

The mere fact that the prior art could be so modified would not have made the modification obvious unless the prior art suggested the desirability of the modification.

221 U.S.P.Q. 1125 (Fed. Cir. 1984).

This doctrine has been routinely embraced by the PTO Board of Appeals. For example, in *Ex parte Kranz*, the Board made it clear that examiner's must identify either an explicit motivation, or a "compelling motivation based upon sound scientific principles." 19 U.S.P.Q.2d 1216, 1218 (BPAI 1990).

Here, the Examiner has merely taken the position that since the genes are known, and some have been placed into a particular plant other than maize, that the introduction of the gene into maize is somehow inherently motivated. It is indeed Appellants' specification that provides for the first time the motivation for introducing the recited genes into maize -- and for this reason it is improper hindsight for the Examiner to rely upon Appellants' own specification to provide the necessary motivation. *In re Deminski*, 230 U.S.P.Q. 313 (Fed. Cir. 1986) This is certainly not the "compelling motivation" based upon "sound scientific principles" referred to by the *Kranz* Board, and certainly an insufficient basis for making out a *prima facie* case of obviousness.

Appellants will now turn to a consideration of each of the individual species of corn/transgene combinations set forth in the

claims pending in the appeal. Appellants have attempted to collect those references referred to in the subject specification and apparently relied upon by the Examiner in completing his basis of rejection. Appellants will address each combination individually, and requests that patentability be considered separately.

- (1) fertile transgenic maize bearing a *bar* gene (claims 129, 135-137)

The gene encoding phosphinothricin acetyl transferase (the *bar* gene, e.g., *Streptomyces hygroscopicus* or *S. viridochromogenes*) inactivates the active ingredient of the herbicide bialophos, phosphinothricin. As reported in Appellants' specification at page 14, lines 17-35, the *bar* gene has been successfully cloned and introduced into other plant species. At page 14, line 35, Appellants refer to the two articles of De Block *et al.* (see Exhibits L and M) as exemplary of articles demonstrating the successful introduction of a *bar* gene into plant species other than maize. In particular, the De Block *et al.* article disclose the introduction of a *bar* gene into tobacco, potato and tomato plants (Exhibit L) and Brassica species (Exhibit M). None of the foregoing plants are monocots, and it is submitted that successful introduction into these plants would in no way suggest that it was possible to introduce the same gene into corn, or what would be the effect of the introduced gene(s) on the resultant corn plant. See, *In re Goodman*, 29 U.S.P.Q.2d 2010 (Fed. Cir. 1993).

Appellants have reviewed each of the two De Block articles and have been unable to find any disclosure in either that would teach

or suggest the introduction of a *bar* gene into maize. Thus, it is difficult to see how such a reference could provide a motivation for introducing the gene into corn. Moreover, even if the Board were to conclude that there was somehow an "implied" motivation, there appears to be no disclosure in either article that would teach how one would proceed to introduce the *bar* gene into corn. The only technique disclosed in either reference is that of *Agrobacterium*-mediated transformation, which for the reasons discussed above with respect to Goldman is submitted to be insufficient and inappropriate in the case of corn.

It is further pointed out that neither of the two De Brock references teach that a *bar* gene can be successfully introduced into corn, and each fail to teach or suggest that a *bar* gene would function appropriately to impart herbicide resistance upon introduction into corn.

- (2) fertile transgenic maize bearing a nitrilase gene (claims 130, 138)

Appellants' specification refers to the nitrilase gene at page 13, lines 2-3, and incorporates by reference the article of Stalker et al. (Exhibit N). As can be seen, the Stalker et al. article relates simply to the cloning of the nitrilase *bxn* gene, and fails to in any way teach or suggest to idea of introducing the gene into the corn genome. Furthermore, the article fails to provide any discussion relevant to how the ordinarily skilled worker would proceed to introduce the gene into corn, fails to provide any evidence that such an introduction would be successful if

attempted, and fails to provide any evidence that the *bxn* would function appropriately if introduced into corn.

In fact, that only discussion in Stalker *et al.* relevant to plants can be found at the very end, where it is mentioned that recent studies have shown that the gene has been introduced into "transgenic plants," but the kind of plants are not identified. A later article by Stalker indicates that these experiments were carried out merely in tobacco, again using *Agrobacterium*-mediated transformation (which is a routine task in dicots such as tobacco).

- (3) fertile transgenic maize bearing a  $\beta$ -lactamase gene (claims 131, 139)

The  $\beta$ -lactamase gene is disclosed in the specification at page 13, lines 15-16, which refers to the article of Sutcliffe (Exhibit 0). The Sutcliffe article relates to the cloning of the ampicillin resistance gene in an *E. coli* host using the pBR322 vector. No disclosure from Sutcliffe relevant to the cloning of the ampicillin resistance gene in corn has been identified, and this article fails to in any way teach or suggest the idea of introducing the gene into the corn genome. Furthermore, the article fails to provide any discussion relevant to how one would proceed to introduce the gene into corn, fails to provide any evidence that such an introduction would be successful if attempted, and fails to provide any evidence that the gene would function appropriately if introduced into corn.



- (4) fertile transgenic maize bearing a *xylE* gene (claims 132, 140)

The *xylE* gene is disclosed in the specification at page 13, line 21, which refers to the article of Zukowski et al. (Exhibit P). The Zukowski article relates to the cloning of the *xylE* gene in *E. coli* and *B. subtilis* hosts. No disclosure from Zukowski relevant to the introduction of the *xylE* gene into corn has been identified, and this article fails to in any way teach or suggest to idea of introducing the gene into the corn genome. Furthermore, the article fails to disclose or suggest how one would proceed to introduce the gene into corn, fails to provide any evidence that such an introduction would be successful if attempted, and fails to provide any evidence that the gene would function appropriately if introduced into corn.

- (5) fertile transgenic maize bearing a galactosidase gene (claims 133, 141)

While the subject specification mentions the use of the galactosidase gene at page 13, line 24, it does not refer to a specific reference. However, Appellants enclose the exemplary reference of Teeri et al. (Exhibit Q) which discloses the introduction of the galactosidase gene into transgenic tobacco. However, as with the previous references, Appellants have been unable to identify any teaching from Teeri relevant to the cloning of the galactosidase gene in corn, and this article fails to in any way teach or suggest to idea of introducing the gene into the corn

genome. Furthermore, the article fails to disclose or suggest how one would proceed to introduce the gene into corn, fails to provide any evidence that such an introduction would be successful if attempted, and fails to provide any evidence that the gene would function appropriately if introduced into corn.

- (6) fertile transgenic maize bearing a tyrosinase gene (claims 134, 142)

The tyrosinase gene is disclosed in the specification at page 13, line 21, which refers to the article of Katz *et al.* (Exhibit R). The Katz article relates to the cloning of the tyrosinase gene. No disclosure from Katz relevant to the cloning of the tyrosinase gene in corn has been identified, and this article fails to in any way teach or suggest to idea of introducing the gene into the corn genome. Furthermore, the article fails to disclose or suggest how one would proceed to introduce the gene into corn, fails to provide any evidence that such an introduction would be successful if attempted, and fails to provide any evidence that the gene would function appropriately if introduced into corn.

- (7) fertile transgenic maize bearing a wheat germ agglutinin gene (claims 103, 143)

Lastly, while the subject specification mentions the use of the wheat germ agglutinin gene at page 17, line 2, it does not refer to a specific reference for the cloning of the gene. However, Appellants enclose the exemplary reference of Raikhel *et al.* (Exhibit S) which discloses the cloning of the wheat germ

agglutinin gene in *E. coli*. However, as with the previous references, Appellants have been unable to identify any teaching from Raikhel et al. directed to the introduction of the wheat germ agglutinin gene in corn, and this article fails to in any way teach or suggest to idea of introducing the gene into the corn genome. Furthermore, the article fails to teach or suggest how one would proceed to introduce the gene into corn, fails to provide any evidence that such an introduction would be successful if attempted, and fails to provide any evidence that the gene would function appropriately if introduced into corn.

**AT BEST, IT CAN MERELY BE SAID THAT THE PRESENT INVENTION IS "OBVIOUS TO TRY" -- AN IMPROPER BASIS FOR CONCLUDING OBVIOUSNESS**

At best, it can merely be said that introduction of the known genes was "obvious to try." Appellants would direct the Board's attention to the case of *In re O'Farrell*, 7 U.S.P.Q.2d 1673, 1680 (Fed. Cir. 1988), which held that in order for a reference or references to obviate an invention, it must be shown that the reference(s) contains:

- (1) detailed enabling methodology for practicing the claimed invention;
- (2) a suggestion for modifying the prior art to practice the claimed invention; and
- (3) evidence suggesting that the invention would be successful.

It is submitted that the present references relied upon by the

Examiner clearly fail to satisfy this tripartite test of *O'Farrell*. In particular, for the reasons discussed above none of the references provides a suggestion for combining the teachings of, e.g., references teaching particular genes with the Goldman reference, and none provides a reasonable expectation that such a combination would be successful, in that for the reasons detailed above there is clear uncertainty as to the operability of the Goldman technique. Moreover, there is certainly no suggestion for modifying these prior disclosures in a manner that would allow one to arrive at the invention, and certainly no evidence that any particular modification would be successful.

In *In re Vaeck*, 20 U.S.P.Q. 1438 (Fed. Cir. 1991), the Federal Circuit took the *O'Farrell* doctrine a step further. In *Vaeck* the Federal Circuit stated that in order for an examiner to make out a *prima facie* case of obviousness two things must be shown: 1) that the prior art would have suggested to those of ordinary skill in the art that they should make the claimed composition; and 2) that the prior art must demonstrate a reasonable expectation of success of the invention. Both the suggestion and the reasonable expectation of success must be founded in the prior art, not in the applicant's disclosure. Here, for the reasons discussed above, we have neither.

**IN THAT COMPOSITIONS ARE CLAIMED, EVEN THE EXISTENCE OF AN OPERABLE METHOD IS INSUFFICIENT TO SUPPORT AN OBVIOUSNESS REJECTION**

The recent case of *In re Deuel*, 34 U.S.P.Q.2d 1210 (Fed. Cir. 1995) provides still further strong support for a conclusion of non-obviousness. In *Deuel*, the Court held that with respect to claims directed to compositions, the fact that methods for making the composition were known or obvious, and even the existence of a motivation to do so, were irrelevant to the question of obviousness of the composition. In order to find the composition obvious, an examiner is required to demonstrate structural obviousness of the composition without reliance upon the existence of an obvious method for its preparation:

The PTO's focus on known methods for potentially isolating the claimed DNA molecules is also misplaced because the claims at issue define compounds, not methods. See *In re Bell*, 991 F.2d 781, 785, 26 USPQ2d 1529, 1532 (Fed. Cir. 1993). In *Bell*, the PTO asserted a rejection based upon the combination of a primary reference disclosing a protein (and its complete amino acid sequence) with a secondary reference describing a general method of gene cloning. We reversed the rejection, holding in part that "[t]he PTO's focus on Bell's method is misplaced. Bell does not claim a method. Bell claims compositions, and the issue is the obviousness of the claimed composition, not of the method by which they are made." *Id.* . . . Thus, even if, as the examiner stated, the existence of general cloning techniques, coupled with knowledge of a protein's structure, might have provided motivation to prepare a cDNA or made it obvious to prepare a cDNA, that does not necessarily make obvious a particular claimed cDNA. "Obvious to try" has long been held not to constitute obviousness. *In re O'Farrell*, 853 F.2d 894, 903, 7 USPQ2d 1673, 1680-81 (Fed. Cir. 1988). (emphasis in original)

34 U.S.P.Q.2d at 1215-16.

It is submitted that the Examiner has failed to meet the burden set by *Deuel* and its predecessors in making out and

supporting a case for structural obviousness of the claimed fertile, transgenic plants. The mere fact that Goldman may disclose techniques that may or may not be useful in the preparation of transgenic maize plants is submitted to be irrelevant under *Deuel*. The *Deuel* and *Bell* line of cases require the Examiner to make a *prima facie* obviousness rejection without reliance upon "general cloning techniques," as admonished by the *Deuel* court. This has not been accomplished in the present case.

***APPELLANTS' BURDEN OF PROOF IS MERELY BY A "PREPONDERANCE OF THE EVIDENCE", NOT "CLEAR AND CONVINCING" AS ARGUED BY THE EXAMINER***

Although the Examiner concluded that "sufficient, clear, and convincing" evidence must be provided that the present invention would not result from the routine use of the Goldman et al. procedures with other known genes, this is not the appropriate burden for Appellants to meet. It is well settled that the proper test of a description in a publication relied on as § 102 prior art requires a determination of whether or not one of ordinary skill in the art would take the disclosure in the publication and combine it with his own knowledge of the art, and from the knowledge be put in possession of the claimed invention. However, the burden placed on Appellants by this type of rejection is to rebut this "presumption of operability" of a reference by a preponderance of the evidence. *In re Sasse*, 207 USPQ 107 (CCPA 1980).

***STRONG SECONDARY EVIDENCE EXISTS IN SUPPORT OF NON-OBVIOUSNESS  
OF THE PRESENT INVENTION***

Appellants would like to review for the Board what is considered to be very strong secondary evidence of non-obviousness, in the form of various scientific articles evidencing the long felt but unresolved need for the invention, the failure of others to achieve the invention, and scientific accolades of peers in the research community for the ultimate achievement of the present invention.

Applicants will turn first to two review articles of Dr. Ingo Potrykus, a well known expert in the field of plant transformation (exhibits T and U). In the former of the foregoing review articles, Dr. Potrykus, as late as June 1990, refers to the serious difficulties associated with genetic engineering of monocots:

...my personal experience in working towards the genetic engineering of cereals for the last 18 years convinces me that we still have serious problems in front of us.... It seems to me that we are really not yet close to such a situation.

p. 535, col. 1-2.

Pessimism pervades Dr. Potrykus' 1989 review article:

Despite considerable efforts in the genetic engineering of plants, and notably achievements in that some species, the world's major cereal crops are proving remarkably recalcitrant to genetic transformation.

Exhibit U, p. 269, abstract.

An article from the journal *Science*, published shortly after the announcement that transgenic, fertile corn has been achieved, emphasizes the long-felt but unresolved need, as well as the failure of others such as Carol Rhodes ("Corn Transformed", Exhibit

V). The article begins by noting that the achievement of fertile, transgenic corn is "the capstone of almost a decade's efforts to genetically engineer this country's most important crop," and then continues by noting the "years of frustration: and a renewed effort to genetically engineer corn begun by Carol Rhodes and her colleagues. The article notes that while Dr. Rhodes and her group were successful in regenerating transformed corn, their "celebrations were short-lived: the resulting plants were infertile."

The *Science* article then refers to attempts by the CIBA-Geigy group, who were said to have achieved the regeneration of non-transgenic corn cells into fertile plants. However, it is pointed out that "these techniques, so far, have not worked with genetically transformed corn."

The Rhodes et al. (see Exhibit W) referred to in the *Science* article discussed above is particularly relevant. The Rhodes article -- which is dated well after the Goldman filing date -- demonstrated using a later technique that while various genes could readily be introduced into corn, of 38 transformed plants derived from 10 different cell lines, NONE were found to be fertile (Exhibit W p. 206, col. 3). The Rhodes article is submitted as evidence of nonobviousness -- it represents the failure of others to achieve fertile, transgenic corn, and demonstrates that the technical breakthrough of the present invention is not the mere introduction of DNA into the corn genome, but the ability to do so in a reproducible manner that achieves fertile, transgenic plants



and transgenic offspring plants. It will be recalled that Rhodes et al. were routinely successful using an electric field to introduce individual genes into corn cells, but all of the plants that were prepared turned out to be infertile. This is a strong secondary consideration that must be considered by the Board in resolving the obviousness question.

In the November 1990 issue of *Bioworld* (Exhibit X), in an article entitled "They Make it Happen in Biotech," Dr. Catherine Mackey, head of the DeKalb transformation research team, was picked as one of four scientists shaping the biotech industry's growth (p. 36). The article refers to the achievement of genetically-engineered corn as "one of the Holy Grails" of monocot transformation. The article then states that monocots such as corn "have been the toughest nuts to crack" in agricultural biotechnology.

Similarly, in the March 1990 issue of *Agricultural Genetics Report*, the race for corn transformation is characterized as biotechnology's "run for the roses" and "Holy Grail":

The Bottom Line: Well, more heard from in agricultural biotechnology's run for the roses -- corn transformation. As we have observed earlier in this space, stable transformation of maize has always been the Holy Grail of agricultural biotechnology.

Exhibit Y, p. 2.

In the April, 1990, issue of the Case Reports of the Ct. Academy of Science and Engineering (Exhibit Z), it is stated that the "DeKalb plants are the first documented fertile, transgenic corn, the result of nearly 10 years of genetic engineering.

The Oct. 1990 issue of the ARI Newsletter (Exhibit AA) observed that the previous report by Sandoz Crop Protection was able to produce only infertile transgenic corn plants.

In the July, 1990, issue of the *Ag Consultant*, an article characterizes the DeKalb achievement as a "major breakthrough in biotechnical research" that "could possibly open up whole new areas of plant improvement." (Exhibit BB).

The October 1990 issue of *Genetic Technology News* included a special section detailing the "breakthrough" represented by genetically-engineered corn. The article lauds the efforts of the DeKalb plant genetics research team and states that "now we know for sure that it is possible to genetically engineer corn" (Exhibit CC).

Further, the *AG Biotechnology News* characterizes the development of this technology as "revolutionary" (Exhibit DD, col. 1).

The DeKalb team's efforts in achieving genetically engineered corn did not go unnoticed in the lay press either. *The Investor's Daily* characterized this achievement as "an advance that other scientists hailed yesterday as a breakthrough" (Exhibit EE). The *Rockford Register Star* characterized this achievement as "the launching point for a genetic engineering revolution" (Exhibit FF, col. 1). The Chicago Tribune quoted the Assistant director of the Washington University Center for Plant Science as the achievement being a "tremendous breakthrough." (Exhibit GG). Moreover, the *Wall Street Journal* recognized the great achievement represented by the

achievement of fertile, transgenic corn:

While scientists have been able to get new genes into monocot plants, the plants previously have ended up sterile and unable to pass the new genes on in their seed. (emphasis supplied)

Exhibit HH, col. 1.

In conclusion, it is submitted that the foregoing strong evidence of non-obviousness conclusively demonstrates the significant achievement represented by the present invention to others in the corn industry. From this evidence it is submitted that a conclusion of non-obviousness is inescapable. Therefore, for each of the foregoing reasons, Appellants respectfully submit that the Board must conclude that the invention defined by the present claims is patentable over the art.

The Board is requested to reconsider and overrule the Examiner's rejection based upon Goldman, taken alone or in combination with the art cited in Appellants' specification.